



Universidade de Aveiro Departamento de Matemática  
2010

**ANABELA VIRGÍNIA  
DOS SANTOS  
FLORES DA ROCHA**

**ESTIMAÇÃO ROBUSTA EM MODELOS LINEARES  
DE EQUAÇÕES SIMULTÂNEAS**



Universidade de Aveiro Departamento de Matemática  
2010

**ANABELA VIRGÍNIA  
DOS SANTOS  
FLORES DA ROCHA**

**ESTIMAÇÃO ROBUSTA EM MODELOS LINEARES  
DE EQUAÇÕES SIMULTÂNEAS**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Matemática, realizada sob a orientação científica da Doutora Maria Manuela Souto de Miranda, Professora Auxiliar do Departamento de Matemática da Universidade de Aveiro e co-orientação do Doutor João António Branco, Professor Associado do Instituto Superior Técnico da Universidade Técnica de Lisboa.

Dedico este trabalho aos meus filhos.

## **o júri**

presidente

**Prof. Doutor Aníbal Guimarães da Costa**  
Professor Catedrático da Universidade de Aveiro.

**Prof<sup>a</sup>. Doutora Maria Manuela Magalhães Hill**  
Professora Catedrática do Instituto Superior de Ciências do Trabalho e da Empresa do Instituto  
Universitário de Lisboa.

**Prof. Doutor João António Branco**  
Professor Associado do Instituto Superior Técnico da Universidade Técnica de Lisboa (Co-  
orientador).

**Prof<sup>a</sup>. Doutora Ana Maria Pires Parente**  
Professora Auxiliar com Agregação do Instituto Superior Técnico da Universidade Técnica de  
Lisboa.

**Prof<sup>a</sup>. Doutora Maria Manuela Souto Miranda**  
Professora Auxiliar da Universidade de Aveiro (Orientadora).

**Prof<sup>a</sup>. Doutora Isabel Maria Simões Pereira**  
Professora Auxiliar da Universidade de Aveiro.

## **agradecimentos**

Queria agradecer aos meus orientadores, à Professora Manuela Souto e ao Professor João Branco, pela possibilidade que me deram de aprender mais, trabalhando com ambos. A compreensão e o apoio que me dispensaram foram essenciais para a realização do trabalho.

Ao Tozé dedico um agradecimento especial, por estar sempre comigo.

Para a minha família e os meus amigos, que sempre me apoiaram, o meu obrigada. Um agradecimento particular para a Tinó, por tudo.

## palavras-chave

Modelo de Equações Simultâneas, Método dos Momentos Generalizado, Estimador Ortogonalizado de Gnanadesikan-Kettenring, Estimação Robusta.

## resumo

Os Modelos de Equações Simultâneas (*SEM*) são modelos estatísticos com muita tradição em estudos de Econometria, uma vez que permitem representar e estudar uma vasta gama de processos económicos. Os estimadores mais usados em *SEM* resultam da aplicação do Método dos Mínimos Quadrados ou do Método da Máxima Verosimilhança, os quais não são robustos. Em Maronna e Yohai (1997), os autores propõem formas de “robustificar” esses estimadores. Um outro método de estimação com interesse nestes modelos é o Método dos Momentos Generalizado (*GMM*), o qual também conduz a estimadores não robustos. Estimadores que sofrem de falta de robustez são muito inconvenientes uma vez que podem conduzir a resultados enganadores quando são violadas as hipóteses subjacentes ao modelo assumido. Os estimadores robustos são de grande valor, em particular quando os modelos em estudo são complexos, como é o caso dos *SEM*. O principal objectivo desta investigação foi o de procurar tais estimadores tendo-se construído um estimador robusto a que se deu o nome de *GMMOGK*. Trata-se de uma versão robusta do estimador *GMM*. Para avaliar o desempenho do novo estimador foi feito um adequado estudo de simulação e foi também feita a aplicação do estimador a um conjunto de dados reais. O estimador robusto tem um bom desempenho nos modelos heterocedásticos considerados e, nessas condições, comporta-se melhor do que os estimadores não robustos usados no estudo. Contudo, quando a análise é feita em cada equação separadamente, a especificidade de cada equação individual e a estrutura de dependência do sistema são dois aspectos que influenciam o desempenho do estimador, tal como acontece com os estimadores usuais. Para enquadrar a investigação, o texto inclui uma revisão de aspectos essenciais dos *SEM*, o seu papel em Econometria, os principais métodos de estimação, com particular ênfase no *GMM*, e uma curta introdução à estimação robusta.

**keywords**

Simultaneous Equations Model, Generalized Method of Moments, Orthogonalized Gnanadesikan-Kettenring Estimator, Robust Estimation.

**abstract**

Simultaneous Equations Models (*SEM*) are statistical models with great tradition in Econometrics since they allow the representation and the study of a large variety of economic processes. The *SEM* estimators are usually obtained by the methods of Least Squares or Maximum Likelihood, which are non-robust estimators. Maronna e Yohai (1997) give robust versions of these estimators. Another non-robust estimator is given by the Generalized Method of Moments (*GMM*). Estimators that suffer from lack of robustness are very inconvenient since they can lead to misleading results when the hypotheses underlying the model are violated. Robust estimators are of great value, in particular when the models under study are complex as *SEM*. The main objective of this research was to search for such estimators and a robust estimator, named *GMMOGK*, was constructed. It is a robust version of *GMM*. To evaluate the performance of this new estimator a convenient simulation study was conducted and the application of the estimator to a real set of data was also carried out. The robust estimator performs well and behaves better than the non-robust estimators used in the study for heteroscedastic models. However the specificity of each individual equation and the dependence structure of the system are two aspects that influence the performance when each equation is analyzed separately. Nevertheless that disadvantage also occurs with the usual estimators.

To frame the research the text includes a review of the essential aspects of *SEM*, their role in econometrics, the main methods of estimation, with particular emphasis to *GMM*, and a short introduction to robust estimation.

# Índice

Agradecimentos	iv
Resumo	v
Abstract	vi
Índice	vii
Lista de Tabelas	x
Lista de Figuras	xii
<b>1 Introdução</b>	<b>1</b>
1.1 Enquadramento do Modelo de Equações Simultâneas . . . . .	1
1.1.1 A Economia . . . . .	2
1.1.2 A Econometria . . . . .	3
1.1.3 A Estatística . . . . .	6
1.2 Objectivos e organização do trabalho . . . . .	7
<b>2 Modelo Linear de Equações Simultâneas</b>	<b>13</b>
2.1 Generalidades . . . . .	14
2.2 Formalização do modelo . . . . .	21
2.2.1 Forma Estrutural . . . . .	21
2.2.2 Forma Reduzida . . . . .	27
2.3 Identificabilidade . . . . .	33
<b>3 Estimação tradicional dos parâmetros de um SEM</b>	<b>43</b>
3.1 Informação limitada e informação completa . . . . .	44
3.2 Utilização do Método dos Mínimos Quadrados em SEM . . . . .	45
3.3 Estimadores de mínimos quadrados com informação limitada . . . . .	47
3.3.1 Estimador OLS . . . . .	48
3.3.2 Estimador de variáveis instrumentais . . . . .	52
3.3.3 Estimador de mínimos quadrados indirectos . . . . .	53



3.3.4	Estimador de mínimos quadrados generalizados . . . . .	57
3.3.5	Estimador de mínimos quadrados em duas fases . . . . .	60
3.4	Estimadores de mínimos quadrados com informação completa . . . . .	66
3.4.1	Estimador de mínimos quadrados em três fases . . . . .	67
3.5	Comparação das matrizes de covariâncias assintóticas dos estimadores <i>2SLS</i> e <i>3SLS</i> . . . . .	72
3.6	Utilização do Método dos Momentos Generalizado em <i>SEM</i> . . . . .	73
3.7	Estimadores <i>GMM</i> com informação limitada . . . . .	82
3.7.1	Estimador <i>GMMIL</i> . . . . .	82
3.8	Estimadores <i>GMM</i> com informação completa . . . . .	87
3.8.1	Estimador <i>GMMIC</i> . . . . .	87
3.9	Comparação das matrizes de covariâncias assintóticas dos estimadores <i>GMMIL</i> e <i>GMMIC</i> . . . . .	90
3.10	Estudo empírico da eficiência relativa entre estimadores de informação limitada e de informação completa . . . . .	94
3.10.1	Modelo proposto por Kmenta (1997) . . . . .	96
3.10.2	Modelo proposto por Judge <i>et al.</i> (1988) - simulação de dados .	100
<b>4</b>	<b>Introdução à estimação robusta</b>	<b>111</b>
4.1	Conceitos gerais . . . . .	112
4.2	Estimação robusta . . . . .	122
4.2.1	Invariância e equivariância . . . . .	122
4.2.2	Estimadores-M . . . . .	126
4.2.3	Outras classes de estimadores . . . . .	136
4.3	Regressão robusta . . . . .	141
4.3.1	Modelo de regressão e método dos mínimos quadrados . . . . .	142
4.4	Estimação robusta em <i>SEM</i> . . . . .	150
<b>5</b>	<b>Versão robusta do estimador <i>GMM</i></b>	<b>153</b>
5.1	Introdução . . . . .	153
5.2	Obtenção de estimativas iniciais dos resíduos . . . . .	155
5.3	Estimação da matriz de covariâncias $\Omega$ . . . . .	156
5.4	Resolução do problema de minimização . . . . .	159
5.5	Algoritmos para o cálculo de estimativas pela versão robusta do <i>GMM</i>	164
5.5.1	Algoritmo HetIC . . . . .	164
5.5.2	Algoritmo HomIC . . . . .	167
<b>6</b>	<b>Estudos de aplicação</b>	<b>169</b>
6.1	Avaliação do desempenho da versão robusta do estimador <i>GMM</i> . . . .	170
6.2	Exemplo proposto por Judge <i>et al.</i> (1988) - dados simulados . . . . .	174

6.3	Exemplo proposto por Maronna e Yohai (1997) - dados reais . . . . .	197
<b>7</b>	<b>Conclusões</b>	<b>203</b>
7.1	Comentários gerais . . . . .	203
7.2	Perspectivas de trabalho futuro . . . . .	208
	<b>Apêndices</b>	<b>211</b>
	<b>Referências Bibliográficas</b>	<b>231</b>

# Lista de Tabelas

2.1	Partição relativa às variáveis que aparecem na equação i . . . . .	36
3.1	Estimadores usuais em <i>SEM</i> . . . . .	45
3.2	Estimativas 2SLS e 3SLS obtidas com systemfit. . . . .	99
3.3	Estimativas 2SLS e 3SLS obtidas com o programa desenvolvido pela autora. . . . .	99
3.4	Estimativas GMMIL e GMMIC obtidas com o programa desenvolvido pela autora. . . . .	100
3.5	Estimativas dos parâmetros estruturais obtidas com T=20 observações, usando variáveis exógenas e matrizes $\Gamma$ , B e $\Sigma$ de acordo com Judge e variáveis endógenas simuladas com erros $N3(0, \Sigma)$ . . . . .	107
3.6	Valores das diferenças entre os determinantes das estimativas das matrizes de covariâncias dos estimadores de informação limitada e completa para as diferentes amostras. . . . .	107
6.1	Médias dos indicadores - caso heterocedástico - valores do GMMOGK obtidos com o algoritmo HetIC. . . . .	182
6.2	Médias dos indicadores - caso homocedástico - valores do GMMOGK obtidos com o algoritmo HetIC. . . . .	182
6.3	Largura das bandas - caso heterocedástico - valores do GMMOGK obtidos com o algoritmo HetIC. . . . .	185
6.4	Largura das bandas - caso homocedástico - valores do GMMOGK obtidos com o algoritmo HetIC. . . . .	185
6.5	Médias dos indicadores - caso homocedástico - valores do GMMOGK obtidos com o algoritmo HomIC. . . . .	192

6.6	Largura das bandas - caso homocedástico - valores do GMMOGK obtidos com o algoritmo HomIC. . . . .	194
6.7	Largura das bandas - caso T5 - valores do GMMOGK obtidos com o algoritmo HetIC. . . . .	195
6.8	Médias dos indicadores - caso T5 - valores do GMMOGK obtidos com o algoritmo HetIC. . . . .	197
6.9	Largura das bandas obtidas - dados da Argentina. . . . .	202
6.10	Valores dos indicadores - dados da Argentina. . . . .	202

# Lista de Figuras

4.1	Gráficos das funções $\Psi$ e $\rho$ de Huber associadas ao estimador-M de localização. . . . .	129
5.1	Esquema de reamostragem - passos 3 a 5 do Algoritmo HetIC. . . . .	165
5.2	Construção da matriz $\mathbf{P}$ - passo 6 do Algoritmo HetIC. . . . .	165
6.1	Resíduos da equação 1 obtidos com HetIC - caso heterocedástico sem contaminação. . . . .	183
6.2	Resíduos da equação 2 obtidos com HetIC - caso heterocedástico sem contaminação. . . . .	184
6.3	Resíduos da equação 3 obtidos com HetIC - caso heterocedástico sem contaminação. . . . .	185
6.4	Resíduos da equação 1 obtidos com HetIC - caso heterocedástico com 20% de contaminação. . . . .	186
6.5	Resíduos da equação 2 obtidos com HetIC - caso heterocedástico com 20% de contaminação. . . . .	186
6.6	Resíduos da equação 3 obtidos com HetIC - caso heterocedástico com 20% de contaminação. . . . .	187
6.7	Resíduos da equação 1 obtidos por HetIC - caso homocedástico sem contaminação. . . . .	187
6.8	Resíduos da equação 2 obtidos por HetIC - caso homocedástico sem contaminação. . . . .	188
6.9	Resíduos da equação 3 obtidos por HetIC - caso homocedástico sem contaminação. . . . .	188

6.10	Resíduos da equação 1 obtidos por HetIC - caso homocedástico com 20% de contaminação. . . . .	189
6.11	Resíduos da equação 2 obtidos por HetIC - caso homocedástico com 20% de contaminação. . . . .	189
6.12	Resíduos da equação 3 obtidos por HetIC - caso homocedástico com 20% de contaminação. . . . .	190
6.13	Resíduos da equação 1 obtidos por HomIC - caso homocedástico com 20% de contaminação. . . . .	190
6.14	Resíduos da equação 2 obtidos por HomIC - caso homocedástico com 20% de contaminação. . . . .	191
6.15	Resíduos da equação 3 obtidos por HomIC - caso homocedástico com 20% de contaminação. . . . .	191
6.16	Resíduos da equação 1 obtidos com HomIC - caso homocedástico sem contaminação. . . . .	193
6.17	Resíduos da equação 2 obtidos com HomIC - caso homocedástico sem contaminação. . . . .	193
6.18	Resíduos da equação 3 obtidos com HomIC - caso homocedástico sem contaminação. . . . .	194
6.19	Resíduos da equação 1 obtidos com HetIC - caso T5. . . . .	195
6.20	Resíduos da equação 2 obtidos com HetIC - caso T5. . . . .	196
6.21	Resíduos da equação 3 obtidos com HetIC - caso T5. . . . .	196
6.22	Resíduos obtidos por <i>GMMIC</i> e por <i>GMMOGK</i> para a 1 <sup>a</sup> equação - dados da Argentina. . . . .	201
6.23	Resíduos obtidos por <i>GMMIC</i> e por <i>GMMOGK</i> para a 2 <sup>a</sup> equação - dados da Argentina. . . . .	201

# Capítulo 1

## Introdução

### 1.1 Enquadramento do Modelo de Equações Simultâneas

O objectivo desta secção introdutória é a apresentação de aspectos fundamentais dos modelos de equações simultâneas. Este tipo de modelos, indispensável à concretização de muitos estudos econométricos, é essencialmente um instrumento de natureza estatística. Nesta secção faz-se o enquadramento destes sistemas de equações, tendo em conta a sua componente econométrica, resultante da estrutura do problema económico que se pretende resolver, e a sua componente estatística. Colocam-se lado a lado estas três componentes (económica, econométrica e estatística), para tentar compreender quais as respectivas contribuições para a análise econométrica.

Qual o papel que desempenha cada um destes saberes na realização de uma análise econométrica? O breve texto que se segue pretende responder a esta pergunta. O que se descreve constitui uma visão simples, sem os detalhes técnicos que um especialista em economia poderia incluir, e resulta de uma reflexão que se entendeu ser útil fazer para se perceber o contexto em que o modelo de equações simultâneas aparece.

Na segunda secção deste capítulo, descreve-se sumariamente o conteúdo dos capítulos seguintes.

### 1.1.1 A Economia

Os modelos de equações simultâneas são de grande utilidade em Econometria se a sua estrutura e modo de funcionamento forem bem conhecidos. Antes de descrever qual é o seu verdadeiro papel, convém relembrar que na base da Econometria está a Economia, motivando os problemas que se espera virem a ser resolvidos pela Econometria. Num primeiro contacto com esta área, percebe-se que o interesse reside em entender a relação entre variáveis económicas, como, por exemplo: i) a relação entre a procura de um bem de consumo disponível no mercado e o preço do mesmo bem, para usar apenas duas variáveis, ou ii) a relação entre a procura do bem de consumo, a oferta desse bem no mercado, o seu preço e o preço de outros bens com ele relacionados, para usar mais do que duas variáveis.

Mas o economista não se fica por esta simplicidade; a sua preocupação prende-se com o entender a natureza e o funcionamento de sistemas económicos complexos: i) relacionados com quantidades macroeconómicas como, por exemplo, o produto nacional bruto (*PNB*) e a inflação ou, ii) relacionados com actividades sectoriais consideradas individualmente como, por exemplo, a produção e o preço do trigo ou a implicação do decréscimo do preço das chamadas por telemóvel na procura deste meio de comunicação. Ao compreender o sistema económico, o economista espera ser capaz de i) efectuar previsões relativas à evolução do sistema, oferecendo aos vários agentes económicos a oportunidade de poderem interferir no sistema, controlando de certo modo a sua evolução, e de ii) testar teorias económicas que tenham sido propostas para explicar o sistema.

O processo de aprendizagem e estudo do problema passa, em primeiro lugar, por construir um modelo que, embora sendo uma simplificação da realidade económica, deve incluir os aspectos mais relevantes do sistema. Um exemplo de modelo usado frequentemente para fins pedagógicos é a relação que iguala o *PNB* à soma de três parcelas, o consumo, o investimento bruto e os gastos do Estado em bens e serviços. Este modelo simplista do *PNB* inclui apenas uma equação relativa à definição do próprio *PNB* mas o que é comum é encontrar modelos que incluem sistemas com várias equações. Neste caso, introduzir certas hipóteses relativas às componentes do *PNB* (sobre o comportamento do consumo e do investimento, por exemplo), equivale



a acrescentar àquela equação outras equações, e é o sistema de equações que constitui a interpretação matemática da teoria económica sobre a determinação conjunta das variáveis consideradas. O número de relações depende dos objectivos do modelo e do grau de explicação que se deseja. Em geral podem distinguir-se dois tipos de equações: comportamentais, que descrevem o comportamento dos vários agentes económicos e outras que se referem a restrições impostas pelo sistema. Além destas equações, o modelo pode integrar certas definições ou identidades económicas.

Para finalizar esta breve exposição de ideias básicas, aqui introduzidas sem preocupações de realismo económico, pode dizer-se que um modelo económico não é mais do que um conjunto de hipóteses, traduzidas em relações matemáticas exactas, entre variáveis económicas que pretendem descrever o comportamento de sectores da economia ou da própria economia em geral.

### **1.1.2 A Econometria**

A partir do modelo económico é possível deduzir um conjunto de equações que integram não só variáveis observáveis como também uma componente relativa a erros, os quais absorvem os aspectos que o sistema não consegue modelar (má especificação do modelo, variáveis que não interessam ao objectivo do modelo, erros de medição e outras perturbações aleatórias). É esta estrutura que é conhecida como um modelo econométrico. Para completar a descrição do modelo econométrico é essencial referir-se as variáveis observáveis são afectadas de erros de observação e indicar a estrutura e a distribuição de probabilidade dos erros inerentes ao modelo. Resumindo, o processo de construção de um modelo econométrico consiste em: i) definir uma teoria económica, ii) traduzir essa teoria em equações matemáticas exactas (modelo económico) e iii) introduzir erros aleatórios nas relações económicas, consubstanciando um tipo de modelo estatístico (modelo econométrico). Esta última fase permite confrontar o modelo com a realidade (efectuando medições, estimando parâmetros e discutindo a validade do modelo) para averiguar se a teoria é, ou não, compatível com a realidade, isto é, se é ou não plausível aceitar essa teoria.

A partir de um mesmo modelo económico inicial podem geralmente derivar-se vários modelos econométricos. Neste processo encontramos a Economia, a Matemática e a

Estatística. A Econometria integra todos estes saberes mas é diferente de qualquer um deles.

Visto desta maneira, o modelo econométrico é um modelo estatístico em relação ao qual se podem e devem adoptar procedimentos habituais: investigar a validade dos pressupostos em que o modelo assenta, estimar os parâmetros, testar hipóteses de interesse (suscitadas pelo próprio modelo ou pela realidade que ele pretende descrever) e usar o modelo assim construído de forma sustentada para efectuar previsões, para justificar políticas económicas e para testar teorias económicas. Com base nestas considerações pode concluir-se que a Econometria tem por objectivo a formulação, a estimação, o teste de modelos econométricos e, finalmente, o seu uso na resolução de problemas económicos. Há muitas outras definições de Econometria, que têm vindo a surgir ao longo do tempo. Haavelmo (1944) dá esta definição:

*"The method of econometric research aims, essentially, at a conjunction of economic theory and actual measurements, using the theory and technique of statistical inference as a bridge pier."*

Tintner (1953) indica outras definições, ainda numa altura em que a Econometria estava no início do seu desenvolvimento. De facto, embora as origens da Econometria se possam localizar nos princípios do século XX (a *Econometric Society* foi fundada em 1930), a econometria formal, como hoje se conhece, começa por volta de 1940. Desde então o desenvolvimento da área tem sido considerável. Mais problemas, mais investigadores e um poder computacional crescente nas últimas décadas justificam o grande e sofisticado desenvolvimento da Econometria. As consequências do progresso científico em Econometria têm sido de grande utilidade para a compreensão da economia das sociedades modernas. Vários especialistas em Econometria viram o resultado do seu trabalho, concretizado em novas metodologias, ser galardoado com o prémio Nobel para as ciências económicas. Dois desses cientistas, cuja obra está directamente ligada ao tema deste trabalho, são: Lawrence Klein (prémio Nobel em 1980) *"for the creation of econometric models and the application to the analysis of economic fluctuations and economic policies"* e Trygve Haavelmo (prémio Nobel em 1989) *"for his clarification on the probability theory foundations of econometrics and his analyses of simultaneous*

*economic structures*”.

O desenvolvimento da Econometria fez crescer o interesse por este último tema, que actualmente é ensinado e investigado em todas as universidades que têm um departamento de economia. A literatura é hoje extremamente volumosa, com um grande número de livros publicados e inúmeros artigos em várias revistas, entre as quais estão revistas especializadas como, por exemplo, *Econometrica*, *the Econometric Reviews*, *Econometric Theory*, *the Journal of Applied Econometrics* e *the Journal of Econometrics*. Em Portugal, a publicação em língua portuguesa, por autores portugueses, parece muito escassa. Na pesquisa efectuada encontrámos dois textos claramente do âmbito deste trabalho, Magalhães (1994), assim como uma tese de mestrado, Penalva (2003) e ainda duas referências relativas a manuais para ensino, Andrade (1993) e Píñheiro e Galego (1999). Destaca-se ainda um trabalho de índole histórica, Machado e Santos Silva (2002), que se refere ao ensino da Econometria em Portugal desde que a disciplina de Econometria apareceu pela primeira vez num curso de Economia de uma universidade portuguesa, o Instituto Superior de Ciências Económicas e Financeiras da Universidade Técnica de Lisboa, no ano lectivo de 1952/53.

Apesar do sucesso que registou nas últimas décadas, a Econometria não conseguiu calar críticas provenientes de vários quadrantes, possivelmente porque é um tema complexo com aspectos polémicos, não esquecendo as suas limitações no que respeita às fragilidades de certas teorias económicas e à natureza não experimental dos dados económicos. Dois dos vários críticos que publicamente se pronunciaram são Leamer (1983) e Hendry (1980). Mais recentemente, o debate incide sobre os papéis da econometria teórica e da econometria aplicada. Aqui, as críticas residem no facto da econometria teórica ter vindo a tornar-se cada vez mais abstracta, esquecendo a preocupação e motivação para a sua utilização prática.

Todos aqueles que estão hoje preocupados com este problema fundamental, consideram também que o principal desafio para a Econometria neste século XXI será o de eliminar o afastamento entre a econometria teórica e a econometria aplicada. E há fundadas esperanças que isso venha a acontecer: as facilidades computacionais existem para proporcionar a produção, a divulgação e o acesso a dados reais; estes meios ficarão mais disponíveis na sociedade, em particular ao nível do próprio ensino e

em revistas científicas, permitindo que estudantes trabalhem em projectos e problemas com contexto da vida real. Por outro lado, lembrando que o método econométrico, essencial em Economia, é também partilhado por outras áreas das ciências sociais (Psicologia, Sociologia e outras), é natural esperar que a complexidade de problemas futuros venha a precisar de novos modelos teóricos para explicar os comportamentos sociais nessas áreas. Isto fará com que a abordagem econométrica venha a precisar de desenvolvimentos inovadores. Krishnakumar (2007) adianta que outra expectativa é um entendimento unificado destas disciplinas, aparentemente separadas, mas partilhando os princípios do método econométrico na resolução dos seus próprios problemas.

### 1.1.3 A Estatística

Ao introduzir perturbações aleatórias (erros aleatórios) no modelo económico teórico, a descrição matemática exacta deste modelo dá lugar a uma descrição estatística das mesmas leis e, a partir daí, a análise do modelo tem uma componente estatística forte que se manifesta no trabalho estatístico habitual, sobretudo em termos de estimação e testes de hipóteses estatísticas. Contudo, não se pode esquecer que esta operação não é simplesmente a operação típica da estatística matemática usada no tratamento de dados experimentais. Os fenómenos económicos não são em geral provenientes da actividade experimental; por exemplo, não é possível fazer variar o preço de um bem, fixando o preço de outros bens com ele relacionados, porque a variação do preço do primeiro arrastará inexoravelmente a variação do preço dos outros. A aplicação directa dos métodos estatísticos clássicos não é adequada, por isso, o uso dos métodos estatísticos tem de ser adaptado à realidade económica.

Outra intervenção da estatística neste processo é ao nível da recolha, organização, descrição e interpretação de dados económicos. Esta actividade, conhecida por estatística económica, consiste sobretudo numa descrição de certos aspectos quantitativos da economia, tentando investigar a possível existência de relações entre variáveis económicas. Nesse caso, não existe um modelo e por isso não há a preocupação da estimação de parâmetros. Porém, esta estatística descritiva pode ser decisiva na concepção de modelos econométricos, pois pode dar pistas sobre a estrutura do fenómeno aleatório e sobre as relações entre as variáveis que ele integra.

Voltando ao modelo econométrico, recorde-se que no mundo económico real se encontram fenómenos descritos por uma só equação, como é o caso dos conhecidos modelos de regressão linear simples ou múltipla e de séries temporais. Mas, a teoria económica ensina que as variáveis associadas a uma determinada teoria variam continuamente e simultaneamente, pelo que cada equação individual que se considere está em geral inserida num sistema de equações, cada uma das quais diz respeito a uma parte da economia que descreve a dependência conjunta dessas variáveis. Quais as implicações da existência de outras equações para além da equação central que se pretende estudar? Deve isolar-se esta equação esquecendo as outras? Deve estudar-se o sistema completo, realizando a análise separada de cada equação do sistema ou a análise conjunta de todas as equações? Os problemas económicos são influenciados por interacções produzidas por muitos agentes económicos e, como tal, parece natural efectuar o estudo conjunto das equações do sistema devido à existência de interacções entre as equações. Não é correcto estudar cada equação isoladamente como se fosse independente das outras. Em vez de equações individuais, deve ser considerado o sistema de equações simultâneas. É este o tema central deste trabalho e que vai ser tratado em vários capítulos da forma que a seguir se indica. A importância do modelo de equações simultâneas em Econometria está bem definida em Hausman (1983, p. 392): “*The simultaneous equation model is perhaps the most remarkable development in econometrics*”.

## 1.2 Objectivos e organização do trabalho

Na presente secção descrevem-se os principais objectivos do trabalho e o modo como foi organizado o texto.

Enquanto modelo estatístico, o Modelo de Equações Simultâneas (*SEM* do inglês *Simultaneous Equations Model*) possui algumas especificidades, aliás decorrentes da sua utilidade como modelo econométrico. Ao ser caracterizado por um sistema de equações algébricas contendo um grande número de parâmetros estatísticos desconhecidos, qualquer estudo de natureza inferencial é necessariamente antecedido por um estudo algébrico, que garanta as condições em que os parâmetros do modelo podem ser estimados. Deste modo, surge a questão da identificabilidade do modelo. Supondo

que são válidas as condições de identificabilidade dos parâmetros do modelo, o passo seguinte, fundamental do ponto de vista estatístico, é o da estimação.

Os métodos de estimação mais divulgados neste tipo de modelos são o dos Mínimos Quadrados e o dos Momentos Generalizado. O Método dos Momentos Generalizado (*GMM* do inglês *Generalized Method of Moments*) foi inicialmente proposto por Hansen (1982) e generaliza o Método dos Momentos ao estimar, por esta via, a esperança de funções de parâmetros que traduzem condições do modelo em estudo. Este método permite estimar parâmetros em condições bastante gerais, nomeadamente, quando os erros são heterocedásticos ou quando o número de condições de momentos é superior ao número de parâmetros a estimar.

Tendo em conta que um *SEM* é definido por um sistema de equações, os métodos de estimação nestes modelos podem dividir-se em dois grupos, os Métodos de Informação Limitada - em que a estimação dos parâmetros é feita separadamente para cada equação - e os Métodos de Informação Completa - em que os parâmetros de todas as equações do modelo são estimados simultaneamente. Existe alguma controvérsia em relação à estimação simultânea ou individual das equações, mas está provado (ver Judge *et al.*, 1988) que usando o princípio dos mínimos quadrados, o método de informação completa é assintoticamente tão ou mais eficiente do que o método de informação limitada. No entanto, alguns autores (por exemplo, Greene, 2003 e Baltagi, 2008) consideram que, em amostras de dimensão moderada, nem sempre se observa a superioridade comprovada teoricamente. Para o caso do método dos momentos generalizado, estudou-se esta questão que resultou em conclusões idênticas às estabelecidas para o método dos mínimos quadrados.

Os métodos de estimação que se referiram até ao momento, resultam em estimadores com boas propriedades quando os erros são normalmente distribuídos. No entanto, não possuem robustez estatística. Considera-se que um método estatístico é robusto quando é pouco sensível a ligeiros afastamentos das hipóteses assumidas para o modelo (pressupostos do modelo). A falta de robustez de um estimador pode originar estimativas dos parâmetros completamente distorcidas. Daí decorre a necessidade de desenvolver métodos robustos que constituam alternativas aos métodos tradicionais. Desde os anos 80 foram desenvolvidas técnicas de estimação robusta para modelos de localização, de

escala e de regressão.

Na sequência do crescente reconhecimento da importância da robustez estatística, enquanto propriedade fundamental de um estimador, têm vindo a ser investigados métodos robustos para modelos e problemas específicos. No caso particular do *SEM*, têm sido feitas várias tentativas, tendo em vista a obtenção de estimadores robustos dos parâmetros destes modelos. Realçam-se as contribuições de Amemiya (1982), de Maronna e Yohai (1997) e Krishnakumar e Ronchetti (1997). No entanto, os estimadores de *GMM* dos parâmetros não têm sido estudados sob o ponto de vista da robustez. Assim, considerou-se ser do maior interesse o desenvolvimento de uma versão robusta do *GMM*, investigação essa que passou a constituir o principal objectivo do estudo da presente tese. Com este objectivo, e dada a diversidade de metodologias de estimação neste tipo de modelos, nomeadamente, a utilização de informação completa ou de informação limitada, estudaram-se essas questões e direccionou-se a investigação da alternativa robusta para a melhor das vias.

Assim, a presente dissertação encontra-se organizada como se passa a descrever.

Para além do presente capítulo, onde se faz a introdução do tema, a tese contém mais seis capítulos.

No segundo capítulo faz-se a apresentação dos Modelos de Equações Simultâneas, inicialmente de um modo informal, de forma a destacar o seu interesse e a familiarizar o leitor com situações em que possam ser aplicados. De seguida, apresenta-se a caracterização do modelo, recordando as suas duas principais representações, a Forma Estrutural e a Forma Reduzida. Por fim, trata-se da questão da identificabilidade. Como se referiu, esta questão tem um interesse particular nestes modelos, e o seu estudo é indispensável, antes de prosseguir com a pesquisa de novos métodos de estimação.

No Capítulo 3 são expostos os métodos tradicionais de estimação dos *SEM*. Descrevem-se dois dos principais métodos de estimação, o Método dos Mínimos Quadrados e o Método dos Momentos Generalizado; em primeiro lugar, explica-se o princípio em que se baseia cada um dos métodos e, de seguida, clarifica-se a forma de aplicação aos *SEM*. O capítulo continua com a apresentação dos estimadores obtidos pelos diferentes métodos (quer por informação completa, quer por informação limitada) e as respectivas

propriedades. Para ilustrar a aplicação dos diversos estimadores e avaliar a superioridade dos métodos de informação completa sobre os de informação limitada em estudos práticos (em dimensão finita), utilizaram-se dois exemplos de modelos econométricos. Os estudos de simulação desenvolvidos num dos exemplos ajudam a explicar os motivos que poderão estar na base das dúvidas referidas em relação à superioridade dos métodos de informação completa.

O Capítulo 4 trata da estimação robusta. Os métodos estatísticos de maior divulgação, nomeadamente os que mais se utilizam para realizar testes de hipóteses, determinar intervalos de confiança ou estimar parâmetros, são usados devido às suas propriedades de optimalidade quando se verifica um conjunto de pressupostos. Perante a falha de tais pressupostos, esses métodos podem conduzir a resultados desastrosos. Relativamente a este tópico, resumem-se os principais conceitos associados à robustez, referindo-se em particular a questão da regressão robusta, a sua importância e as principais formas de implementar esta metodologia. Referem-se ainda as principais abordagens publicadas sobre estimação robusta em *SEM*.

No quinto capítulo apresenta-se a versão robusta do estimador *GMM*, proposta neste trabalho, que foi designada por estimador *GMMOGK* (estimador do método dos momentos generalizado, ortogonalizado de *Gnanadesikan - Kettenring*). Esta proposta é desenvolvida essencialmente em três fases: i) na primeira fase são determinadas estimativas iniciais dos parâmetros, aplicando um estimador robusto - estimador dos mínimos quadrados aparados (*LTS*), uma vez que as propriedades dos estimadores usados no cálculo das estimativas iniciais vão ser determinantes para a qualidade das estimativas finais, ii) a segunda fase consiste em estimar de forma robusta a matriz de covariâncias dos erros do modelo usando métodos recentes de robustificação de escala, iii) com a robustez sempre em vista, a terceira fase envolve a minimização da função de Huber aplicada aos resíduos, depois de devidamente ponderados com estimativas obtidas em ii). Este procedimento conduz à determinação de um estimador global dos parâmetros, que é robusto por construção.

No sexto capítulo efectua-se uma avaliação do desempenho do estimador proposto, a partir de dois estudos de aplicação - um baseado em dados simulados e outro em dados reais. Efectuaram-se estudos empíricos para proceder a uma avaliação prática



do estimador *GMMOGK*. Usaram-se dados reais da economia argentina, já estudados e analisados na literatura (Maronna e Yohai (1997)), de forma a facilitar a comparação dos resultados e a aferir o interesse do método agora proposto. Adicionalmente fizeram-se estudos de simulação, sob diversos cenários: alterando a dimensão da amostra, a percentagem de contaminação, a distribuição dos erros (normal ou t de Student, homocedástica ou heterocedástica).

Por fim, no sétimo capítulo, referem-se os principais resultados conseguidos, as conclusões que daí decorrem e as perspectivas do trabalho futuro.

A tese contém ainda informação adicional, organizada numa secção de apêndices, onde se apresentam alguns cálculos e demonstrações que suportam os restantes capítulos.

## Capítulo 2

# Modelo Linear de Equações Simultâneas

Este capítulo tem como objectivo fazer uma apresentação resumida do Modelo Linear de Equações Simultâneas e das principais questões que se colocam na sua análise e na sua utilização. Este tema encontra-se tratado sob a forma de um ou mais capítulos em livros de Econometria publicados por diversos autores, como por exemplo Schmidt (1976), Judge *et al.* (1988), Johnston e DiNardo (1997), o livro que marcou o ensino da econometria em todo o mundo, desde a sua primeira edição em 1963, Kmenta (1997), Mittelhammer *et al.* (2000), Greene (2003), Gujarati (2003), Baltagi (2008) e Wooldridge (2009). Em português, poucos trabalhos têm sido publicados por portugueses sobre o tratamento estatístico destes modelos, havendo apenas a referir Magalhães (1994) e Penalva (2003).

Uma vez que na presente tese apenas serão tratados modelos de estrutura linear, usar-se-á a designação *SEM* ao longo do trabalho, para referir estes modelos, sem estar constantemente a fazer alusão à linearidade. O capítulo encontra-se organizado em três secções. Na primeira secção, procede-se à apresentação informal do *SEM* tendo em vista destacar o seu interesse e familiarizar o leitor com situações que justificam a sua aplicação, tendo-se usado para ilustração vários exemplos, entre os quais o Modelo de Keynes, tão famoso entre economistas. A segunda secção compreende a descrição da notação matricial a usar e ainda a apresentação formal do modelo, relembrando as suas duas principais representações, nomeadamente, a Forma Estrutural e a Forma

Reduzida. A última secção trata da identificabilidade do sistema. Esta questão é particularmente importante nestes modelos pois, sendo caracterizados por sistemas de equações lineares, torna-se essencial avaliar a estrutura algébrica subjacente antes de prosseguir com o seu estudo estatístico.

## 2.1 Generalidades

O Modelo de Equações Simultâneas é um modelo estatístico caracterizado por um sistema de equações que explicitam relações de dependência entre variáveis, assumindo simultaneamente a interdependência entre equações (à frente, neste trabalho, explicar-se-á melhor esta noção de interdependência). Um *SEM* diz-se **linear** se for linear em relação às variáveis e em relação aos parâmetros.

Um *SEM* pode compreender vários tipos de equações. Algumas traduzem relações fundamentais entre variáveis relativas ao fenómeno que está a ser modelado. Estas relações, chamadas **equações estruturais**, estabelecem ligações entre variáveis. A aleatoriedade do modelo fica contemplada introduzindo nestas equações um termo aditivo aleatório, usualmente designado por erro do modelo. Para além dessas variáveis, as equações estruturais contêm também parâmetros desconhecidos, correspondendo estes últimos aos coeficientes das diversas variáveis; estes parâmetros são chamados **parâmetros estruturais** do modelo. Nas aplicações econométricas, as equações estruturais descrevem a estrutura das relações entre as variáveis económicas, relações essas que são fornecidas pela própria teoria económica.

As restantes equações do sistema também traduzem ligações entre as variáveis, mas não envolvem termos relativos a erros aleatórios, nem parâmetros desconhecidos. Em aplicações económicas, estas equações formalizam muitas vezes identidades contabilísticas ou condições de equilíbrio.

Os exemplos que se passam a apresentar procuram esclarecer a diferença entre os dois tipos de equações, ao mesmo tempo que permitem começar a introduzir a terminologia específica destes modelos. Os aspectos relativos a detalhes e complexidades de natureza económica e à dinâmica do modelo não serão discutidos, pois não é esse o objectivo deste estudo.

*Exemplo 2.1.1.* (Modelo de Keynes)

Este primeiro exemplo é um caso paradigmático da literatura, pela importância que teve no desenvolvimento de estudos macroeconómicos. Trata-se de um modelo Keynesiano que descreve a interdependência entre o consumo e o rendimento. Este modelo, de acordo com Judge *et al.* (1988), é definido pelo sistema,

$$\begin{cases} y_t = c_t + x_t \\ c_t = \beta + \gamma y_t + \varepsilon_t \end{cases}, \quad (2.1.1)$$

onde as variáveis aleatórias  $c_t$  e  $y_t$  representam, respectivamente, o consumo e o rendimento no período  $t$ ; a variável  $x_t$  representa o investimento no período  $t$  e  $\varepsilon_t$  é o termo relativo ao erro aleatório do modelo. No modelo existem dois parâmetros desconhecidos,  $\beta$  e  $\gamma$ , havendo especial interesse em  $\gamma$ , pelo seu significado económico. Este parâmetro é designado por propensão marginal ao consumo e representa o montante adicional de consumo quando o rendimento sofre um aumento unitário, sabendo-se pela teoria económica que  $0 < \gamma < 1$ .

Enquanto que, de acordo com a primeira equação,  $y_t$  depende de  $c_t$ , pela segunda equação resulta que  $c_t$  também vai depender de  $y_t$ . Desde já se chama a atenção para a relação de interdependência que se verifica entre as variáveis  $c_t$  e  $y_t$  - estas variáveis são designadas por variáveis endógenas. Pelo contrário, o valor do investimento  $x_t$  é pré-determinado autonomamente em relação ao sistema, pelo que esta variável se designa por variável exógena.

Para identificar quais são as equações estruturais do modelo, note-se que a segunda equação contém um termo relativo ao erro aleatório e dois parâmetros desconhecidos; trata-se, portanto, de uma equação estrutural. Na primeira equação isso não acontece. A equação corresponde a uma identidade económica que, neste caso, traduz uma condição de equilíbrio (rendimento = consumo + investimento).  $\diamond$

*Exemplo 2.1.2.* (Modelo simplificado da oferta e da procura)

Exemplos muito comuns de sistemas de equações simultâneas são os modelos da oferta e da procura. O caso particular do "Modelo simplificado da oferta e da procura" é muito popular na literatura, podendo ser visto, por exemplo, em Kmenta (1997). De acordo com a teoria económica, sabe-se que, num mercado em equilíbrio, o preço de

um bem e a quantidade transaccionada desse bem são determinados pela intersecção das curvas da procura e da oferta. Admitindo que as funções da procura e da oferta são lineares, pode escrever-se o modelo do seguinte modo

$$\begin{cases} Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 Y_t + u_{1t} \\ Q_t = \beta_0 + \beta_1 P_t + u_{2t} \end{cases},$$

onde, ao fim do período  $t$ ,  $Q_t$  representa a quantidade transaccionada do bem no mercado em equilíbrio,  $P_t$  indica o preço de equilíbrio desse bem e  $Y_t$  traduz o rendimento dos consumidores. À semelhança do *Exemplo 2.1.1.*, também aqui há duas variáveis que são determinadas conjuntamente, as variáveis  $Q_t$  e  $P_t$ , que, por isso, são chamadas variáveis endógenas. A variável  $Y_t$  é exógena e  $u_{1t}$  e  $u_{2t}$  são os erros aleatórios. A primeira equação do sistema traduz a função da procura e a segunda equação representa a função da oferta. As constantes  $\alpha_0, \alpha_1, \alpha_2, \beta_0$  e  $\beta_1$  são os parâmetros do modelo que, de acordo com a teoria económica, verificam as condições  $\alpha_1 \leq 0, \alpha_2 \geq 0$  e  $\beta_1 \geq 0$ . Note-se ainda que, ao estudar a relação entre quantidade e preço, não é possível considerar apenas uma das equações isoladamente, pois ambas as equações são necessárias para determinar as variáveis endógenas, concluindo-se assim que se trata de um modelo de equações simultâneas com duas equações estruturais.

De acordo com a teoria económica, a oferta e a procura descrevem e explicam a variação do preço e da quantidade de bens vendidos em mercados plenamente competitivos. A procura representa a quantidade de um bem que um consumidor estaria disposto a comprar a um dado preço, num determinado período de tempo. A lei da procura defende que, em geral, quanto maior for o preço, menor será a quantidade de bem procurada. Por outro lado, a oferta representa a quantidade desse bem que um produtor está disposto a colocar no mercado a um dado preço, no mesmo período de tempo. A lei da oferta estabelece que quanto maior for o preço do bem, maior será a quantidade desse bem que os produtores irão querer produzir. O preço e a quantidade de equilíbrio serão encontrados quando a quantidade produzida é igual à quantidade consumida, ou seja, quando o mercado está em situação de equilíbrio.  $\diamond$

#### *Exemplo 2.1.3.* (Modelo de Klein I)

Um outro exemplo típico de um modelo econométrico que é descrito por um *SEM* é

o modelo de Klein I. Trata-se de um modelo macroeconómico proposto por Klein, em 1950, para descrever a economia dos EUA no período entre as duas guerras mundiais. O desenvolvimento de modelos econométricos deve muito a Laurence R. Klein, laureado com o prémio Nobel da economia em 1980, precisamente pela sua extraordinária contribuição para a "criação de modelos econométricos e sua aplicação à análise económica das flutuações e da política económica", como já tinha sido mencionado no capítulo 1. Em 1950 publicou um livro (Klein, 1950) onde apresenta três modelos Keynesianos, pensados para a economia americana relativa ao período de 1921 a 1941. O mais simples destes três modelos é conhecido por modelo de Klein I, modelo que agora aqui se descreve. Possivelmente devido à sua simplicidade interpretativa e facilidade de tratamento, o modelo é muito atractivo pelo seu sucesso pedagógico em explicar a estrutura e funcionamento dos SEM, em particular a questão central da estimação dos parâmetros. Não admira, portanto, que muitos livros incluam a descrição e estudo do modelo de Klein I nos seus textos. É o caso, por exemplo de Goldberger (1964), Christ (1966), Theil (1971), Desai (1976), Intriligator *et al.* (1996), Gujarati (2003) e Greene (2003). De acordo com Greene (2003), o modelo é definido por um sistema de seis equações - as primeiras três representam, respectivamente, a função do consumo, a função do investimento e a procura de mão-de-obra, medida pelo salário pago pelo sector privado, e traduzem o comportamento dos respectivos agentes económicos (consumidores, investidores, salários). As restantes equações traduzem identidades económicas, estabelecidas de acordo com a legislação em vigor, que completam e identificam o sistema, tornando-o susceptível de estimação. O modelo de Klein I é definido pelo sistema:

$$\left\{ \begin{array}{lcl} C_t & = & \alpha_0 + \alpha_1 P_t + \alpha_2 P_{t-1} + \alpha_3 (W_t^p + W_t^g) + u_{1t} \\ I_t & = & \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + \beta_3 K_{t-1} + u_{2t} \\ W_t^p & = & \gamma_0 + \gamma_1 A_t + \gamma_2 X_t + \gamma_3 X_{t-1} + u_{3t} \\ X_t & = & C_t + I_t + G_t \\ P_t & = & X_t - T_t - W_t^p \\ K_t & = & K_{t-1} + I_t \end{array} \right. ,$$

onde,

- $C_t$  representa o consumo no ano  $t$ ;
- $I_t$  mede o investimento no ano  $t$ ;

- $G_t$  refere-se à despesa não salarial do governo no ano  $t$ ;
- $P_t$  corresponde aos lucros do sector privado no ano  $t$ ;
- $W_t^p$  traduz o salário pago pelo sector privado no ano  $t$ ;
- $W_t^g$  traduz o salário pago pelo governo no ano  $t$ ;
- $K_t$  representa o capital social no final do ano  $t$ ;
- $X_t$  refere-se ao produto do sector privado no ano  $t$ ;
- $T_t$  corresponde ao imposto cobrado às empresas privadas no ano  $t$ ;
- $A_t = t - 1931$  traduz a tendência temporal no ano  $t$ , medida em anos desde 1931.

O modelo contém seis variáveis endógenas, nomeadamente,  $C_t, I_t, W_t^p, X_t, P_t$  e  $K_t$ , uma vez que os valores destas variáveis são determinados pelo sistema. As variáveis  $P_{t-1}, K_{t-1}$  e  $X_{t-1}$  são variáveis desfasadas no tempo; por isso são tratadas como exógenas, uma vez que no ano  $t$  os seus valores já são conhecidos. O sistema inclui ainda as variáveis exógenas  $W_t^g, G_t, T_t$  e  $A_t$ , cujos valores são determinados exteriormente ao sistema. O modelo contém os parâmetros  $\alpha_j, \beta_j$  e  $\gamma_j, j = 0, \dots, 3$  e os erros aleatórios  $u_{1t}, u_{2t}$  e  $u_{3t}$ . Do ponto de vista estatístico, o interesse incide sobre a estimação dos parâmetros contidos nas três primeiras equações. Klein estimou os parâmetros usando dados de 1921 a 1941, os quais se encontram disponíveis no programa *R* e no site <http://pages.stern.nyu.edu/~wgreene/text/econometricanalysis.htm> da sexta edição (2008) do texto online (Table F 13.1).

O modelo de Klein I é também um modelo dinâmico pois contém coeficientes autoregressivos, ou seja, certos coeficientes referem-se às variáveis endógenas relativas a diferentes momentos no tempo. ◇

Tal como acontece nos exemplos anteriores, no caso geral, um *SEM* inclui dois tipos de variáveis, as variáveis endógenas e as variáveis exógenas, cujas definições se apresentam de seguida.

**Definição 2.1.1. Variável endógena** é aquela cujos valores são determinados conjuntamente com outras variáveis dentro do sistema e que, portanto, depende dos valores de outras variáveis no sistema. □

**Definição 2.1.2. Variável exógena** é aquela cujos valores são determinados exteriormente ao sistema, mas que influencia os valores das variáveis endógenas. O modelo nada diz sobre o modo como estas variáveis são determinadas.  $\square$

De um modo geral, designam-se por variáveis predeterminadas todas as variáveis exógenas, exógenas desfasadas e endógenas desfasadas. A distinção entre variáveis endógenas e variáveis exógenas não é um tema pacífico em econometria e, como tal, o assunto tem levantado certa polémica que pode ser comprovada em relatos encontrados em várias publicações como, por exemplo, Engle *et al.* (1983) e Leamer (1985), entre outros. As definições que aqui se apresentam acompanham a linha tradicional que é geralmente seguida e abrange um grande consenso.

Ao contrário do que acontece com outros modelos, num *SEM* algumas equações envolvem mais do que uma variável dependente ou endógena. Para determinar os valores destas variáveis são necessárias tantas equações quantas as variáveis endógenas envolvidas no sistema. Surge assim a noção de sistema completo.

**Definição 2.1.3.** O **sistema** de equações diz-se **completo** quando o número de equações é igual ao número de variáveis endógenas.  $\square$

Quando esta condição não se verificar, diz-se que o sistema não está completo, o que deverá ser consequência de alguma falha na modelação do fenómeno em estudo.

Uma breve revisão da literatura estatística e econométrica mostra que existe alguma falta de consenso em relação à designação do modelo, pelo que se justifica a apresentação de algumas considerações sobre o assunto, antes da sua caracterização formal. Em particular, é especialmente importante estabelecer a distinção entre o modelo em estudo e outros modelos semelhantes, nomeadamente, o modelo de equações estruturais e o Modelo de Regressão Linear Multivariado (*MRLM*).

De acordo com Kmenta (1997), um *SEM* designa um modelo estatístico formalizado através de um sistema de equações simultâneas, em que todas as equações do sistema são necessárias para determinar o valor de pelo menos uma variável endógena incluída do modelo. Note-se que esta caracterização do modelo implica que pelo menos uma das equações inclua mais do que uma variável endógena; pelo que a definição



de Kmenta constitui um critério de modelação baseado na relação entre o número de equações e o número de variáveis endógenas. Observe-se que, ignorando as condições distribucionais, a formalização do *SEM* pelo sistema pode dificultar a distinção entre um *SEM* e um *MRLM*. No entanto, a noção de variável endógena já torna implícito no sistema as condições de dependência entre variáveis. De facto, verifica-se que num *SEM*, uma variável endógena pode ser variável dependente numa equação e aparecer como variável explicativa, ou regressor, noutra equação. Consequentemente, essa variável endógena, que é regressor numa equação, está frequentemente correlacionada com o erro aleatório da mesma equação.

Por outro lado, um *MRLM* também é estabelecido a partir de um conjunto de equações entre variáveis, as quais incluem parâmetros e termos relativos a erros aleatórios. No entanto, os erros nas diversas equações de um *MRLM* não são correlacionados; para além disso, o *MRLM* assume que os regressores são determinísticos, ou que, sendo aleatórios, não estão correlacionados com os erros. Estas condições não ocorrem num *SEM*.

A distinção entre modelo de equações simultâneas e modelo de equações estruturais já é menos clara na literatura. Alguns autores usam indiferentemente a mesma terminologia para designar ambos os modelos. A confusão é ampliada pelo facto da sigla *SEM* abreviar as duas designações em inglês (*Simultaneous Equations Models* e *Structural Equations Models*). Na tentativa de esclarecer a diferença seguiu-se Kmenta (1997); e note-se, desde já, que os modelos de equações simultâneas podem ser considerados como uma versão simplificada dos modelos de equações estruturais. Uma característica importante destes últimos modelos é o facto de envolverem variáveis não observáveis, designadas por variáveis latentes. As variáveis latentes são variáveis que, pela sua natureza, não são observáveis de forma directa, pelo que a informação que se tem sobre elas resulta da medição dos seus efeitos noutras variáveis (indicadores) que, estas sim, podem ser directamente observadas. Um modelo de equações estruturais é formado por duas partes: a primeira é composta por relações estruturais envolvendo variáveis latentes, podendo estas ser exógenas ou endógenas, e a segunda é composta por equações de medição que relacionam as variáveis latentes com os respectivos indicadores. Concluindo, um *SEM* corresponde a uma particularização do modelo de

equações estruturais, em que as equações estruturais não contêm variáveis latentes, ou seja, todas as variáveis endógenas e exógenas são observáveis.

Apesar da base comum destas duas famílias de modelos, na prática, fica-se com a ideia de que o uso dos modelos de equações simultâneas decorre paralelamente ao uso dos modelos de equações estruturais, como se nada houvesse a ligá-los, para além da designação *SEM*, que os econométristas traduzem de uma maneira e os especialistas em ciências sociais traduzem de outra.

## 2.2 Formalização do modelo

Dependendo da forma como são apresentadas as equações do sistema, um *SEM* pode ser caracterizado de diferentes modos. Esta secção é dedicada à apresentação formal das duas principais representações do *SEM*, a Forma Estrutural e a Forma Reduzida.

### 2.2.1 Forma Estrutural

Considere-se um *SEM* constituído por  $M$  equações, tais que a  $t$ -ésima observação do modelo verifica o sistema

$$\begin{cases} \nu_{11}y_{t1} + \dots + \nu_{M1}y_{tM} + \beta_{11}x_{t1} + \dots + \beta_{k1}x_{tk} + e_{t1} & = 0 \\ \nu_{12}y_{t1} + \dots + \nu_{M2}y_{tM} + \beta_{12}x_{t1} + \dots + \beta_{k2}x_{tk} + e_{t2} & = 0 \\ \vdots & \vdots \\ \nu_{1M}y_{t1} + \dots + \nu_{MM}y_{tM} + \beta_{1M}x_{t1} + \dots + \beta_{kM}x_{tk} + e_{tM} & = 0 \end{cases},$$

onde cada equação  $i$  ( $i = 1, \dots, M$ ) contém: um erro aleatório  $e_{ti}$ , as variáveis endógenas  $y_{t1}, \dots, y_{tM}$ , os parâmetros estruturais associados às variáveis endógenas  $\nu_{ji}$  ( $j = 1, \dots, M$ ) e os parâmetros estruturais  $\beta_{ri}$  ( $r = 1, \dots, k$ ) associados às  $k$  variáveis exógenas  $x_{t1}, \dots, x_{tk}$ . Assume-se que o número total de variáveis endógenas coincide com o número  $M$  de equações, ou seja, que o sistema é completo. O índice  $t$  é usado para indexar as observações, sendo  $T$  o número de observações da amostra ( $t = 1, \dots, T$ ). Quando um *SEM* é expresso na forma anterior, diz-se que se encontra na **Forma Estrutural**.

Passando à notação vectorial para representar uma amostra de dimensão  $T$  do modelo, denotem-se por  $\mathbf{y}_i$  o vector constituído pelas  $T$  observações da  $i$ -ésima variável endógena ( $i = 1, \dots, M$ ), por  $\mathbf{x}_j$  o vector constituído pelas  $T$  observações da  $j$ -ésima

variável exógena ( $j = 1, \dots, k$ ) e por  $\mathbf{e}_i$  o vector de dimensão  $T$  dos erros aleatórios associados à  $i$ -ésima equação ( $i = 1, \dots, M$ ). Assim, a amostra verifica o sistema

$$\begin{cases} \nu_{11}\mathbf{y}_1 + \dots + \nu_{M1}\mathbf{y}_M + \beta_{11}\mathbf{x}_1 + \dots + \beta_{k1}\mathbf{x}_k + \mathbf{e}_1 & = \mathbf{0} \\ \nu_{12}\mathbf{y}_1 + \dots + \nu_{M2}\mathbf{y}_M + \beta_{12}\mathbf{x}_1 + \dots + \beta_{k2}\mathbf{x}_k + \mathbf{e}_2 & = \mathbf{0} \\ \vdots & \vdots \\ \nu_{1M}\mathbf{y}_1 + \dots + \nu_{MM}\mathbf{y}_M + \beta_{1M}\mathbf{x}_1 + \dots + \beta_{kM}\mathbf{x}_k + \mathbf{e}_M & = \mathbf{0} \end{cases} \quad (2.2.1)$$

Reescrevendo (2.2.1) na forma matricial, obtém-se a **Forma Estrutural matricial** do *SEM*, que para o conjunto das  $T$  observações, vem

$$\underbrace{\mathbf{Y}\mathbf{\Gamma} + \mathbf{X}\mathbf{B} + \mathbf{E}}_{(T \times M)(M \times M) + (T \times k)(k \times M) + (T \times M) = (T \times M)} = \mathbf{0}, \quad (2.2.2)$$

onde  $\mathbf{Y}$  é a matriz das observações das variáveis endógenas,  $\mathbf{X}$  é a matriz das observações das variáveis exógenas e  $\mathbf{E}$  é a matriz dos erros aleatórios, ou seja,

$$\begin{aligned} \mathbf{Y}_{(T \times M)} &= [y_{ti}]_{t=1, \dots, T, i=1, \dots, M}, \\ \mathbf{X}_{(T \times k)} &= [x_{tj}]_{t=1, \dots, T, j=1, \dots, k}, \\ \mathbf{E}_{(T \times M)} &= [e_{ti}]_{t=1, \dots, T, i=1, \dots, M}. \end{aligned}$$

Estas matrizes podem ser representadas noutras forma, de acordo com o Apêndice A.1.1. As matrizes  $\mathbf{\Gamma}$  e  $\mathbf{B}$  contêm os parâmetros estruturais do modelo, correspondendo, respectivamente, aos coeficientes das variáveis endógenas e aos coeficientes das variáveis exógenas, isto é,

$$\mathbf{\Gamma}_{(M \times M)} = [\nu_{ji}] \text{ e } \mathbf{B}_{(k \times M)} = [\beta_{ri}], i, j = 1, \dots, M, r = 1, \dots, k.$$

A partir de (2.2.2) é possível exprimir o sistema, para cada observação  $t$ , em termos das matrizes dos coeficientes. Com efeito, mostra-se no Apêndice A.1.2 (a) que

$$\mathbf{Y}'_t \mathbf{\Gamma} + \mathbf{x}'_t \mathbf{B} + \mathbf{e}'_t = \mathbf{0}_{(1 \times M)}, t = 1, \dots, T,$$

onde  $\mathbf{Y}'_t$ ,  $\mathbf{x}'_t$  e  $\mathbf{e}'_t$  representam a linha  $t$  das matrizes  $\mathbf{Y}$ ,  $\mathbf{X}$  e  $\mathbf{E}$ , ou seja, as transpostas de  $\mathbf{Y}_t$ ,  $\mathbf{x}_t$  e  $\mathbf{e}_t$ , respectivamente.

Por outro lado, fixando a equação  $i$  e considerando o conjunto das  $T$  observações, mostra-se no Apêndice A.1.2 (b), que essa equação se pode escrever como

$$\mathbf{Y}\mathbf{\Gamma}_i + \mathbf{X}\mathbf{B}_i + \mathbf{e}_i = \mathbf{0}_{(T \times 1)}, i = 1, \dots, M, \quad (2.2.3)$$

onde  $\Gamma_i$ ,  $\mathbf{B}_i$  e  $\mathbf{e}_i$  representam a coluna  $i$  das matrizes  $\mathbf{\Gamma}$ ,  $\mathbf{B}$  e  $\mathbf{E}$ , respectivamente.

Antes de enunciar as hipóteses assumidas para o modelo, é ainda necessário introduzir mais alguma notação. Designar-se-á por  $\sigma_{it}^2$  ( $i = 1, \dots, M, t = 1, \dots, T$ ) a variância do erro da equação  $i$  para a observação  $t$ , ou seja,  $\sigma_{it}^2 = V(e_{ti}|\mathbf{x}_t)$  e por  $\sigma_{ijt}$  a covariância entre os erros das equações  $i$  e  $j$  para a observação  $t$ , ou seja,  $\sigma_{ijt} = Cov(e_{ti}, e_{tj}|\mathbf{x}_t)$ .

Uma vez que as variáveis  $\mathbf{x}_t$  são exógenas e, portanto, assumem valores previamente determinados, o condicionamento por  $\mathbf{x}_t$  não será evidenciado de modo a não sobrecarregar a notação, já de si pesada, no estudo destes modelos.

Usando a notação vectorial dos erros, com  $\mathbf{e}'_t = \begin{bmatrix} e_{t1} & \dots & e_{tM} \end{bmatrix}$ , representa-se por  $\Sigma_t$  a sua matriz de covariâncias, ou seja,

$$\begin{aligned} \Sigma_t &= E[\mathbf{e}_t \mathbf{e}'_t | \mathbf{x}_t] \\ &= \begin{bmatrix} V(e_{t1}) & \dots & Cov(e_{t1}, e_{tM}) \\ \vdots & \ddots & \vdots \\ Cov(e_{tM}, e_{t1}) & \dots & V(e_{tM}) \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{1t}^2 & \sigma_{12t} & \dots & \sigma_{1Mt} \\ \sigma_{21t} & \sigma_{2t}^2 & \dots & \sigma_{2Mt} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M1t} & \sigma_{M2t} & \dots & \sigma_{Mt}^2 \end{bmatrix}, (t = 1, \dots, T). \end{aligned} \quad (2.2.4)$$

No caso particular do modelo ser homocedástico, as variâncias e as covariâncias são iguais para qualquer  $t$ , isto é,  $\sigma_{it}^2$  e  $\sigma_{ijt}$  não variam com  $t$ , sendo então representados por  $\sigma_i^2$  e  $\sigma_{ij}$ , respectivamente. Consequentemente, a matriz  $\Sigma_t$  definida em (2.2.4) é constante para qualquer observação  $t$  do modelo e passa a ter a forma

$$\Sigma = \Sigma_t = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1M} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M1} & \sigma_{M2} & \dots & \sigma_M^2 \end{bmatrix}.$$

Por outro lado, denota-se por  $\Omega_{ij}$  a matriz das covariâncias dos erros associados às equações  $i$  e  $j$  ( $i, j = 1, \dots, M$ ), ou seja,

$$\Omega_{ij} = E[\mathbf{e}_i \mathbf{e}'_j | \mathbf{X}] = E \begin{bmatrix} e_{1i}e_{1j} & \dots & e_{1i}e_{Tj} \\ \vdots & \ddots & \vdots \\ e_{Ti}e_{1j} & \dots & e_{Ti}e_{Tj} \end{bmatrix}.$$

No caso particular de uma única equação  $i$ , a matriz  $\mathbf{\Omega}_{ii}$  será representada por  $\mathbf{\Omega}_i$ .

Relacionando o modelo na Forma Estrutural com o Modelo de Regressão Linear Multivariado, note-se que um *SEM* pode ser encarado como uma generalização do *MRLM*, na medida em que a equação  $\mathbf{Y} = \mathbf{XB} + \mathbf{E}$ , típica do *MRLM*, corresponde a um *SEM* na forma estrutural (2.2.2), para o caso particular  $\mathbf{\Gamma} = \mathbf{I}_M$ , onde  $\mathbf{I}_M$  é a matriz identidade de ordem  $M$ . Esta particularização tem consequências directas sobre a homocedasticidade dos erros e o não correlacionamento entre erros e regressores (que se verificam no *MRLM*), condições essas que não são impostas no *SEM*.

Para concluir a caracterização do *SEM* na forma estrutural, falta apresentar um conjunto de hipóteses que, em geral, são assumidas sobre as variáveis envolvidas. A apresentação das hipóteses vai sendo acompanhada por alguns comentários, sempre que se considere pertinente.

HFE.1 Os erros são gerados por um processo multivariado estacionário de componentes  $e_{ti}$  não correlacionadas ao longo do tempo, ou seja:

HFE.1.1 Os erros têm valor médio nulo, isto é,

$$E[\mathbf{e}_t | \mathbf{x}_t] = \mathbf{0},$$

onde  $\mathbf{e}_t'$  é a linha  $t$  da matriz  $\mathbf{E}$ , para  $t = 1, \dots, T$ .

HFE.1.2 A matriz de covariâncias dos erros pode variar de observação para observação, ou seja,

$$\mathbf{\Sigma}_t = E[\mathbf{e}_t \mathbf{e}_t' | \mathbf{x}_t], \forall t = 1, \dots, T.$$

Para cada  $t$ , admite-se a existência de correlação entre os erros das diferentes equações.

HFE.1.3 Os erros correspondentes a observações diferentes são não correlacionados, isto é,

$$E[e_{ti} e_{sj} | \mathbf{x}_t, \mathbf{x}_s] = 0, t \neq s; t, s = 1, \dots, T; i, j = 1, \dots, M.$$

Na ausência de correlação entre os erros relativos a observações diferentes, diz-se que não há autocorrelação. Então, para  $t \neq s$ , vem que

$$E[\mathbf{e}_t \mathbf{e}_s' | \mathbf{x}_t, \mathbf{x}_s] = \begin{bmatrix} Cov(e_{t1}, e_{s1}) & \dots & Cov(e_{t1}, e_{sM}) \\ \vdots & \ddots & \vdots \\ Cov(e_{tM}, e_{s1}) & \dots & Cov(e_{tM}, e_{sM}) \end{bmatrix} = \mathbf{0}_{(M \times M)}.$$

Na sequência das hipóteses HFE.1.1 a HFE.1.3, interessa acrescentar os comentários seguintes.

- (i) Note-se que, da conjunção de HFE.1.2 com HFE.1.3, se assume que

$$E[e_{ti} e_{sj} | \mathbf{x}_t, \mathbf{x}_s] = \begin{cases} \begin{cases} \sigma_{ijt}, \text{ se } i \neq j \\ \sigma_{it}^2 = \sigma_{iit}, \text{ se } i = j \end{cases}, & \text{se } t = s \\ 0, & \text{se } t \neq s \end{cases},$$

para  $t, s = 1, \dots, T$  e  $i, j = 1, \dots, M$ .

- (ii) Como consequência de HFE.1, tem-se que para cada equação  $i$ , a matriz de covariâncias  $\mathbf{\Omega}_i$  dos erros é uma matriz diagonal, isto é, que

$$\mathbf{\Omega}_i = E[\mathbf{e}_i \mathbf{e}_i' | \mathbf{X}] = \text{diag} \begin{bmatrix} \sigma_{i1}^2 & \dots & \sigma_{iT}^2 \end{bmatrix}. \quad (2.2.5)$$

- (iii) No caso homocedástico, em que o valor da variância do erro de cada equação  $i$  não varia com  $t$ , isto é, quando  $\sigma_{it}^2 = \sigma_i^2, \forall t$ , esta matriz simplifica-se, tomando a forma,

$$\mathbf{\Omega}_i = E[\mathbf{e}_i \mathbf{e}_i' | \mathbf{X}] = \sigma_i^2 \mathbf{I}_T. \quad (2.2.6)$$

- (iv) De modo análogo, a matriz de covariâncias dos erros correspondentes a diferentes equações também é uma matriz diagonal, isto é,

$$\mathbf{\Omega}_{ij} = E[\mathbf{e}_i \mathbf{e}_j' | \mathbf{X}] = \text{diag} \begin{bmatrix} \sigma_{ij1} & \dots & \sigma_{ijT} \end{bmatrix}. \quad (2.2.7)$$

- (v) No caso homocedástico, em que o valor da covariância entre os erros de diferentes equações não varia com  $t$ , ou seja,  $\sigma_{ijt} = \sigma_{ij}, \forall t$ , esta matriz simplifica-se, tomando a forma

$$\mathbf{\Omega}_{ij} = E[\mathbf{e}_i \mathbf{e}_j' | \mathbf{X}] = \sigma_{ij} \mathbf{I}_T.$$

HFE.2 Para cada  $t$  fixo, a distribuição condicionada do vector de erros  $\mathbf{e}_t$  é uma normal multivariada, ou seja,  $\mathbf{e}_t|\mathbf{x}_t \sim N_M(\mathbf{0}, \Sigma_t)$ ,  $t = 1, \dots, T$ .

Note-se que na literatura é usual assumir esta hipótese com a condição adicional de  $\Sigma_t$  ser constante em relação a  $t$ , ou seja, do modelo ser homocedástico em relação a  $t$ .

HFE.3 A matriz  $\mathbf{X}$  é constituída por  $T$  observações de um vector aleatório de dimensão  $k$  e tem característica  $k$ , o que traduz que as colunas de  $\mathbf{X}$  são linearmente independentes. Como consequência, tem-se que o número de linhas de  $\mathbf{X}_{(T \times k)}$ , ou seja o número de observações  $T$ , deve ser tal que  $T \geq k$ .

HFE.4 Admite-se que a matriz

$$\Sigma_{\mathbf{X}\mathbf{X}(k \times k)} = p \lim_{T \rightarrow \infty} \left[ \frac{\mathbf{X}'\mathbf{X}}{T} \right] \quad (2.2.8)$$

é uma matriz finita e definida positiva, onde  $p \lim_{T \rightarrow \infty}$  representa o limite em probabilidade.

HFE.5 Os elementos de  $\mathbf{X}$  são uniformemente limitados.

HFE.6 O erro aleatório associado à observação  $t$  no conjunto de todas as equações,  $\mathbf{e}_t$ , é não correlacionado com o vector das variáveis exógenas dessa observação,  $\mathbf{x}_t$ ,  $t = 1, \dots, T$ , isto é,

$$E[\mathbf{X}'\mathbf{E}] = p \lim_{T \rightarrow \infty} \left[ \frac{\mathbf{X}'\mathbf{E}}{T} \right] = \mathbf{0}. \quad (2.2.9)$$

Esta hipótese estabelece a ortogonalidade entre as variáveis exógenas e os erros aleatórios, pelo que a condição (2.2.9) é usualmente chamada condição de ortogonalidade. Para além disso, importa referir que a condição (2.2.9) é verificada por variáveis exógenas e não por variáveis endógenas.

HFE.7 A matriz  $\mathbf{\Gamma}$  é uma matriz não singular.

Esta hipótese equivale a admitir que o determinante de  $\mathbf{\Gamma}$  é não nulo, sendo uma condição suficiente para que o modelo seja completo.

A partir das hipóteses assumidas para os erros aleatórios, por serem estacionários e não correlacionados para diferentes observações, resulta, por uma generalização da Lei

Fracção dos Grande Números, que

$$p \lim_{T \rightarrow \infty} \left[ \frac{\mathbf{E}'\mathbf{E}}{T} \right] = E \left[ \frac{\mathbf{E}'\mathbf{E}}{T} | \mathbf{X} \right] = \mathring{\Sigma}, \quad (2.2.10)$$

onde  $\mathring{\Sigma} = \begin{bmatrix} \frac{1}{T} \sum_{t=1}^T \sigma_{1t}^2 & \cdots & \frac{1}{T} \sum_{t=1}^T \sigma_{1Mt} \\ \vdots & \ddots & \vdots \\ \frac{1}{T} \sum_{t=1}^T \sigma_{Mt} & \cdots & \frac{1}{T} \sum_{t=1}^T \sigma_{Mt}^2 \end{bmatrix}$ , de acordo com desenvolvimento no Apêndice

A.1.3. A condição (2.2.10) traduz que a matriz de covariâncias amostrais dos erros da forma estrutural converge para a matriz de covariâncias populacionais dos erros da forma estrutural, a qual será representada por  $\mathring{\Sigma}$ .

Apresenta-se a seguir a Forma Reduzida de um *SEM*. A importância desta forma de representar o modelo reside no facto de permitir um enquadramento e tratamento estatístico do modelo análogo ao do *MRLM*, sendo aplicáveis algumas metodologias e resultados disponíveis para a regressão.

## 2.2.2 Forma Reduzida

O sistema correspondente à forma (2.2.2) é formado por  $M$  equações e depende de  $M$  variáveis endógenas, de variáveis exógenas e de erros aleatórios. De acordo com HFE.7, existe a matriz  $\Gamma^{-1}$ ; logo, é possível exprimir as variáveis endógenas em função das variáveis exógenas e dos erros aleatórios. Assim, partindo do modelo na forma estrutural matricial,  $\mathbf{Y}\Gamma + \mathbf{X}\mathbf{B} + \mathbf{E} = \mathbf{0}$ , pode exprimir-se cada variável endógena como uma função linear dependente apenas de variáveis exógenas e de erros aleatórios. Assim, tem-se que

$$\mathbf{Y} = -\mathbf{X}\mathbf{B}\Gamma^{-1} - \mathbf{E}\Gamma^{-1},$$

e designa-se por **Forma Reduzida** associada ao sistema (2.2.2), a equação

$$\underbrace{\mathbf{Y} = \mathbf{X}\Pi + \mathbf{V}}_{(T \times M) = (T \times k)(k \times M) + (T \times M)}, \quad (2.2.11)$$

onde

$$\Pi_{(k \times M)} = -\mathbf{B}\Gamma^{-1}, \quad (2.2.12)$$

$$\mathbf{V}_{(T \times M)} = -\mathbf{E}\Gamma^{-1}. \quad (2.2.13)$$



As matrizes  $\mathbf{\Pi}$  e  $\mathbf{V}$  contêm, respectivamente, os parâmetros e os erros aleatórios da forma reduzida.

Note-se que a coluna  $i$  de  $\mathbf{\Pi}$  é um vector  $k$ -dimensional que contém os coeficientes da equação  $i$  na forma reduzida e será denotada por  $\boldsymbol{\pi}_i (i = 1, \dots, M)$ . De modo idêntico,  $\mathbf{v}_i$  representa um vector  $T$ -dimensional, correspondente à coluna  $i$  da matriz dos erros  $\mathbf{V} (i = 1, \dots, M)$ .

De seguida apresenta-se um exemplo publicado em Judge *et al.* (1988), com o qual se pretende ilustrar os cálculos necessários para passar da forma estrutural de um *SEM* à forma reduzida.

*Exemplo 2.2.1.* Considere-se o seguinte SEM, definido na forma estrutural por

$$\begin{cases} -\mathbf{y}_1 + \mathbf{y}_2 & + 2\mathbf{x}_2 & + \mathbf{e}_1 & = \mathbf{0} \\ 2\mathbf{y}_1 - \mathbf{y}_2 + 3\mathbf{x}_1 & + \mathbf{x}_3 + \mathbf{e}_2 & = \mathbf{0} \end{cases}.$$

Na forma matricial estrutural correspondente a (2.2.2), o vector das variáveis endógenas é  $\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 \end{bmatrix}$ , o vector das variáveis exógenas é  $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{bmatrix}$  e as matrizes dos respectivos parâmetros são

$$\mathbf{\Gamma} = \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix} \text{ e } \mathbf{B} = \begin{bmatrix} 0 & 3 \\ 2 & 0 \\ 0 & 1 \end{bmatrix}.$$

Para passar o sistema à forma reduzida, calcula-se a inversa da matriz  $\mathbf{\Gamma}$ , que neste caso é  $\mathbf{\Gamma}^{-1} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$ , e substitui-se em  $\mathbf{Y} = -\mathbf{XB}\mathbf{\Gamma}^{-1} - \mathbf{E}\mathbf{\Gamma}^{-1}$ . Explicitando, tem-se que

$$\begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 \end{bmatrix} = - \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{bmatrix} \begin{bmatrix} 0 & 3 \\ 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} - \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}.$$

Logo, o SEM fica na forma reduzida  $\mathbf{Y} = \mathbf{X}\mathbf{\Pi} + \mathbf{V}$ , para

$$\mathbf{\Pi} = \begin{bmatrix} 0 & 3 \\ 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 2 & 4 \\ 1 & 1 \end{bmatrix} \text{ e } \mathbf{V} = - \begin{bmatrix} \mathbf{e}_1 + \mathbf{e}_2 & 2\mathbf{e}_1 + \mathbf{e}_2 \end{bmatrix}.$$

Considerando o sistema com as duas equações, vem

$$\begin{cases} y_1 = -3x_1 - 2x_2 - x_3 - e_1 - e_2 \\ y_2 = -3x_1 - 4x_2 - x_3 - 2e_1 - e_2 \end{cases}.$$

◇

Chama-se a atenção para o facto da forma reduzida (2.2.11) se traduzir por uma equação matricial análoga à que caracteriza o modelo de regressão multivariado - com matriz de variáveis dependentes  $\mathbf{Y}$ , matriz de regressores  $\mathbf{X}$ , matriz de parâmetros  $\mathbf{\Pi}$  e matriz de erros  $\mathbf{V}$ . Como se referiu, a diferença entre ambos os modelos não reside na forma da equação (2.2.11), mas sim nas propriedades distribucionais envolvidas, muito particularmente nas que se referem aos erros e aos regressores.

À semelhança do que acontece com a forma estrutural, também para a forma reduzida há interesse em analisar as distribuições dos vectores linha e dos vectores coluna de (2.2.11), os quais correspondem, respectivamente, a cada observação  $t$  para o conjunto de todas as  $M$  equações e a cada equação  $i$  para o conjunto de todas as  $T$  observações.

A partir da forma (2.2.11), destacando a linha que corresponde à observação  $t$ , pode escrever-se a forma reduzida do modelo,

$$\mathbf{Y}'_t = \mathbf{x}'_t \mathbf{\Pi} + \mathbf{v}'_t, t = 1, \dots, T.$$

Fixando agora a coluna  $i$  da forma (2.2.11), resulta a equação  $i$  do modelo na forma reduzida relativamente às  $T$  observações,

$$\mathbf{y}_i = \mathbf{X} \boldsymbol{\pi}_i + \mathbf{v}_i, i = 1, \dots, M.$$

Tendo em conta que os erros da forma reduzida estão definidos como combinação linear dos erros da forma estrutural (por (2.2.13)), as hipóteses assumidas para os erros  $\mathbf{E}$ , da forma estrutural, vão reflectir-se nos erros  $\mathbf{V}$  da forma reduzida. Assim, considerando que  $\mathbf{e}'_t$  e  $\mathbf{v}'_t$  representam a linha  $t$  das matrizes  $\mathbf{E}$  e  $\mathbf{V}$ , respectivamente, onde  $\mathbf{v}'_t = -\mathbf{e}'_t \mathbf{\Gamma}^{-1}$ , e tendo em conta que  $\mathbf{\Gamma}_{(M \times M)}$  é constante, para cada observação  $t$ , os erros  $\mathbf{v}_t$  verificam as seguintes propriedades ( $t = 1, \dots, T$ ).

PFR.1 Os erros têm esperança nula, pois por HFE.1.1, vem que

$$E[\mathbf{v}_t|\mathbf{x}_t] = E\left[(-\mathbf{e}_t'\mathbf{\Gamma}^{-1})'|\mathbf{x}_t\right] = (\mathbf{\Gamma}^{-1})' E[\mathbf{e}_t|\mathbf{x}_t] = \mathbf{0}.$$

PFR.2 Os erros têm matriz de covariâncias  $\Sigma_t^v = E[\mathbf{v}_t\mathbf{v}_t'|\mathbf{x}_t]$ , a qual, por HFE.1.2, é da forma

$$\begin{aligned}\Sigma_{t(M \times M)}^v &= E[\mathbf{v}_t\mathbf{v}_t'|\mathbf{x}_t] \\ &= (\mathbf{\Gamma}^{-1})' E[\mathbf{e}_t\mathbf{e}_t'|\mathbf{x}_t] \mathbf{\Gamma}^{-1} \\ &= (\mathbf{\Gamma}^{-1})' \Sigma_t \mathbf{\Gamma}^{-1}.\end{aligned}\tag{2.2.14}$$

PFR.3 Os erros são não correlacionados, para diferentes observações, pois, de acordo com HFE.1.3, e para  $t \neq s$ ,

$$E[\mathbf{v}_t\mathbf{v}_s'|\mathbf{x}_t, \mathbf{x}_s] = (\mathbf{\Gamma}^{-1})' E[\mathbf{e}_t\mathbf{e}_s'|\mathbf{x}_t, \mathbf{x}_s] \mathbf{\Gamma}^{-1} = \mathbf{0}.$$

PFR.4 Os erros têm distribuição normal

$$\mathbf{v}_t|\mathbf{x}_t \sim N(\mathbf{0}, \Sigma_t^v), t = 1, \dots, T.$$

Esta propriedade decorre da hipótese HFE.2 e do facto de  $\mathbf{v}_t = -(\mathbf{\Gamma}^{-1})' \mathbf{e}_t$ .

Em termos de distribuição assintótica, podem desde já concluir-se as seguintes propriedades:

PFR.5 A matriz de covariâncias amostrais dos erros da forma reduzida converge para a matriz das covariâncias populacionais dos erros da forma reduzida, a qual será representada por  $\mathring{\Sigma}^v$ .

Pela definição de  $\mathbf{V}$  e por (2.2.10),

$$\begin{aligned}p \lim_{T \rightarrow \infty} \left[ \frac{\mathbf{V}'\mathbf{V}}{T} \right] &= p \lim_{T \rightarrow \infty} \left[ \frac{(-\mathbf{E}\mathbf{\Gamma}^{-1})'(-\mathbf{E}\mathbf{\Gamma}^{-1})}{T} \right] \\ &= p \lim_{T \rightarrow \infty} \left[ \frac{(\mathbf{\Gamma}^{-1})' \mathbf{E}' \mathbf{E} \mathbf{\Gamma}^{-1}}{T} \right] \\ &= (\mathbf{\Gamma}^{-1})' \left[ p \lim_{T \rightarrow \infty} \left( \frac{\mathbf{E}' \mathbf{E}}{T} \right) \right] \mathbf{\Gamma}^{-1} \\ &= (\mathbf{\Gamma}^{-1})' \mathring{\Sigma} \mathbf{\Gamma}^{-1}_{(M \times M)} = \mathring{\Sigma}^v.\end{aligned}\tag{2.2.15}$$

PFR.6 As variáveis exógenas  $\mathbf{X}$  são não correlacionadas com os erros da forma reduzida  $\mathbf{V}$ .

De facto, por HFE.6,

$$\begin{aligned} p \lim_{T \rightarrow \infty} \left[ \frac{\mathbf{X}'\mathbf{V}}{T} \right] &= p \lim_{T \rightarrow \infty} \left[ \frac{\mathbf{X}'(-\mathbf{E}\mathbf{\Gamma}^{-1})}{T} \right] \\ &= p \lim_{T \rightarrow \infty} \left[ \frac{\mathbf{X}'\mathbf{E}}{T} \right] \mathbf{\Gamma}^{-1} = \mathbf{0}_{(k \times M)}. \end{aligned} \quad (2.2.16)$$

Ao assumir a hipótese da distribuição normal, decorrem as proposições seguintes que estabelecem as distribuições assintóticas conjuntas das variáveis exógenas e dos erros, tanto para a forma estrutural como para a forma reduzida; apresenta-se apenas a demonstração da última proposição, uma vez que as restantes podem ser consultados em Schmidt (1976).

Antes de prosseguir, torna-se necessário introduzir mais alguma notação.

- (i) Representa-se vectorialmente uma matriz  $\mathbf{A}$  por  $vec(\mathbf{A})$ , o vector formado pelas colunas de  $\mathbf{A}$ , isto é,

$$\text{se } \mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}_{(m \times n)} \Rightarrow vec(\mathbf{A}) = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_n \end{bmatrix}_{(mn \times 1)}.$$

- (ii) Utiliza-se o símbolo  $\dot{\sim}$  para representar a distribuição assintótica.
- (iii) O produto de Kronecker entre duas matrizes  $\mathbf{A} = [a_{ij}]_{(K1 \times M1)}$  e  $\mathbf{B} = [b_{rs}]_{(K2 \times M2)}$  será denotado por  $\otimes$  e definido, como usualmente, por

$$\mathbf{A} \otimes \mathbf{B} = [a_{ij}\mathbf{B}]_{((K1 \times K2) \times (M1 \times M2))}.$$

**Proposição 2.2.1.** Em relação à forma estrutural de um *SEM*, definida em (2.2.2), tem-se que

$$vec \frac{\mathbf{X}'\mathbf{E}}{\sqrt{T}} | \mathbf{X} \dot{\sim} N \left( \mathbf{0}, \dot{\Sigma} \otimes \Sigma_{\mathbf{xx}} \right).$$

□

**Proposição 2.2.2.** Supondo a hipótese HFE.4, tem-se que

$$vec \frac{\mathbf{X}'\mathbf{V}}{\sqrt{T}} | \mathbf{X} \dot{\sim} N \left( \mathbf{0}, \left( (\Gamma^{-1})' \dot{\Sigma} \Gamma^{-1} \right) \otimes \Sigma_{\mathbf{XX}} \right).$$

□

A distribuição assintótica conjunta das variáveis endógenas, das variáveis exógenas e dos erros da forma reduzida, fica estabelecida na proposição que se segue.

**Proposição 2.2.3.** Para um *SEM* na forma reduzida, tem-se que

$$p \lim_{T \rightarrow \infty} \frac{1}{T} \begin{bmatrix} \mathbf{Y}'\mathbf{Y} & \mathbf{Y}'\mathbf{X} & \mathbf{Y}'\mathbf{V} \\ \mathbf{X}'\mathbf{Y} & \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{V} \\ \mathbf{V}'\mathbf{Y} & \mathbf{V}'\mathbf{X} & \mathbf{V}'\mathbf{V} \end{bmatrix} = \begin{bmatrix} \left( \Pi' \Sigma_{\mathbf{XX}} \Pi + (\Gamma^{-1})' \dot{\Sigma} \Gamma^{-1} \right) & (\Pi' \Sigma_{\mathbf{XX}}) & (\Gamma^{-1})' \dot{\Sigma} \Gamma^{-1} \\ (\Sigma_{\mathbf{XX}} \Pi) & \Sigma_{\mathbf{XX}} & \mathbf{0} \\ (\Gamma^{-1})' \dot{\Sigma} \Gamma^{-1} & \mathbf{0} & (\Gamma^{-1})' \dot{\Sigma} \Gamma^{-1} \end{bmatrix},$$

sendo esta uma matriz quadrada de ordem  $(M + k + M)$ .

*Demonstração.* Expressando  $\mathbf{Y}$  na forma reduzida, vem:

$$p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{Y}'\mathbf{Y}) = p \lim_{T \rightarrow \infty} \frac{1}{T} [(\mathbf{X}\Pi + \mathbf{V})'(\mathbf{X}\Pi + \mathbf{V})].$$

Efectuando o produto indicado e por a matriz  $\Pi$  ser constante, resulta que

$$\begin{aligned} p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{Y}'\mathbf{Y}) &= \Pi' \left[ p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{X}'\mathbf{X}) \right] \Pi + \Pi' \left[ p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{X}'\mathbf{V}) \right] \\ &\quad + \left[ p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{V}'\mathbf{X}) \right] \Pi + p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{V}'\mathbf{V}). \end{aligned}$$

Tendo em conta os resultados (2.2.8), (2.2.15) e (2.2.16), vem

$$\begin{aligned} p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{Y}'\mathbf{Y}) &= \Pi' \Sigma_{\mathbf{XX}} \Pi + \Pi' \mathbf{0} + \mathbf{0} \Pi + (\Gamma^{-1})' \dot{\Sigma} \Gamma^{-1} \\ &= \Pi' \Sigma_{\mathbf{XX}} \Pi + (\Gamma^{-1})' \dot{\Sigma} \Gamma^{-1}. \end{aligned}$$

Para obter a forma do elemento (2, 1) da matriz, o procedimento é idêntico, mas desta vez recorrendo a (2.2.8) e (2.2.16).

$$\begin{aligned} p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{X}'\mathbf{Y}) &= p \lim_{T \rightarrow \infty} \frac{1}{T} [\mathbf{X}'(\mathbf{X}\Pi + \mathbf{V})] \\ &= \left[ p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{X}'\mathbf{X}) \right] \Pi + p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{X}'\mathbf{V}) \\ &= \Sigma_{\mathbf{XX}} \Pi + \mathbf{0}. \end{aligned}$$

É ainda por (2.2.16) que se obtém o elemento (3, 1) da matriz:

$$\begin{aligned}
 p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{V}' \mathbf{Y}) &= p \lim_{T \rightarrow \infty} \frac{1}{T} [\mathbf{V}' (\mathbf{X} \mathbf{\Pi} + \mathbf{V})] \\
 &= \left[ p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{V}' \mathbf{X}) \right] \mathbf{\Pi} + p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{V}' \mathbf{V}) \\
 &= \mathbf{0} \mathbf{\Pi} + (\mathbf{\Gamma}^{-1})' \mathring{\mathbf{\Sigma}} \mathbf{\Gamma}^{-1}.
 \end{aligned}$$

Os restantes elementos da matriz são obtidos por simetria, tendo em conta PFR.6, (2.2.15), (2.2.8) e que a matriz  $(\mathbf{\Gamma}^{-1})' \mathring{\mathbf{\Sigma}} \mathbf{\Gamma}^{-1}$  é simétrica.  $\square$

Para terminar esta secção, note-se que se apresentaram dois modelos estatísticos, definidos pelas formas (2.2.2) e (2.2.11), para representar o mesmo *SEM*. Em ambas as formas, será necessário estimar as matrizes de parâmetros,  $\mathbf{\Gamma}$  e  $\mathbf{B}$  da forma estrutural ou  $\mathbf{\Pi}$  da forma reduzida. No entanto, antes de proceder à estimação dos parâmetros, deve efectuar-se o estudo da identificabilidade do modelo.

## 2.3 Identificabilidade

Um modelo diz-se identificável quando são identificáveis todos os seus parâmetros, no sentido em que podem ser univocamente determinados pela distribuição conjunta das variáveis observáveis. Atendendo à terminologia usual nos *SEM*, a identificabilidade do modelo refere-se à identificabilidade dos parâmetros estruturais do modelo, contidos nas matrizes  $\mathbf{\Gamma}$  e  $\mathbf{B}$ .

De facto, sem informação adicional ou sem informação à priori sobre  $\mathbf{Y}$ ,  $\mathbf{B}$  e  $\mathbf{\Sigma}$ , os parâmetros estruturais do modelo não são identificados. Isto acontece porque, sem restrições à priori, os modelos definidos por

$$\mathbf{Y} \mathbf{\Gamma} + \mathbf{X} \mathbf{B} + \mathbf{E} = \mathbf{0} \text{ e por } \mathbf{Y} \mathbf{\Gamma} \mathbf{A} + \mathbf{X} \mathbf{B} \mathbf{A} + \mathbf{E} \mathbf{A} = \mathbf{0} \quad (2.3.1)$$

conduzem à mesma forma reduzida, para qualquer matriz não singular  $\mathbf{A}_{(M \times M)}$ . Nesta situação, diz-se que os modelos referidos são indistinguíveis em termos da informação amostral.

Considere-se um *SEM* na forma estrutural  $\mathbf{Y} \mathbf{\Gamma} + \mathbf{X} \mathbf{B} + \mathbf{E} = \mathbf{0}$ . Recorrendo às técnicas usuais no *MRLM* para estimar os parâmetros, a matriz dos regressores conteria

variáveis endógenas, o que faria com que os estimadores de mínimos quadrados dos parâmetros contidos em  $\mathbf{\Gamma}$  e  $\mathbf{B}$  não fossem consistentes, no caso geral. No entanto, esse problema fica ultrapassado ao trabalhar com o *SEM* na forma reduzida. De facto, os parâmetros da forma reduzida são sempre identificáveis; no entanto, o mesmo não acontece com os parâmetros da forma estrutural, podendo ser necessária informação adicional para regressar à parametrização inicial do modelo. Assim, o estudo da identificabilidade do modelo deve anteceder o estudo da estimação dos parâmetros da forma estrutural.

Depois de introduzida a ideia geral da identificabilidade dos parâmetros estruturais, apresenta-se uma definição proposta em Schmidt (1976), para o caso particular de um *SEM*.

**Definição 2.3.1.** Um parâmetro do modelo estrutural na forma  $\mathbf{Y}\mathbf{\Gamma} + \mathbf{X}\mathbf{B} + \mathbf{E} = \mathbf{0}$  diz-se **identificado** se e só se o seu valor pode ser deduzido a partir do conhecimento dos parâmetros do modelo na forma reduzida. Se o parâmetro não for identificado, diz-se que é **subidentificado**.  $\square$

**Definição 2.3.2.** Uma equação é **identificada** se todos os parâmetros dessa equação forem identificados.  $\square$

Para estudar a identificabilidade, interessa destacar a equação  $i$  do sistema (2.2.1),

$$\nu_{1i}\mathbf{y}_1 + \dots + \nu_{Mi}\mathbf{y}_M + \beta_{1i}\mathbf{x}_1 + \dots + \beta_{ki}\mathbf{x}_k + \mathbf{e}_i = \mathbf{0} .$$

Embora na forma geral, apresentada em (2.2.1), todas as variáveis do sistema apareçam em cada equação, nas situações práticas, apenas algumas variáveis aparecem em cada equação, ou seja, é frequente que os coeficientes de algumas variáveis sejam nulos.

**Definição 2.3.3.** Chamam-se **restrições de exclusão** a equações do tipo

$$\nu_{ij} = 0 \vee \beta_{ri} = 0, i, j = 1, \dots, M, r = 1, \dots, k,$$

onde  $\nu_{ij}$  e  $\beta_{ri}$  são parâmetros do sistema (2.2.1).  $\square$

As restrições de exclusão correspondem à existência de elementos nulos nas matrizes  $\Gamma$  e  $B$ , o que traduz a ausência de algumas variáveis nalgumas equações.

Por outro lado, interessa também proceder à normalização das equações do sistema.

**Definição 2.3.4.** Chama-se **regra de normalização** ao procedimento que permite reescrever o sistema, de modo a que, em cada equação, se isole uma variável endógena no primeiro membro.  $\square$

Sem perda de generalidade, pode atribuir-se o coeficiente  $-1$  a uma variável endógena em cada equação, o que equivale a multiplicar cada equação por uma constante. Esse procedimento pode traduzir-se através de condições do tipo

$$\nu_{ii} = -1, i = 1, \dots, M.$$

Quando numa equação  $i$ , se procede à sua normalização, fixando  $\nu_{ii} = -1$ , e se admite ainda que algumas variáveis não aparecem nessa equação, isso corresponde a introduzir uma partição nos vectores de observações, nos vectores de parâmetros e nos vectores dos erros, de modo a isolar em blocos as variáveis afectadas por restrições de exclusão. Deste modo, será associado o símbolo  $(*)$  às variáveis excluídas e utilizar-se-á a seguinte notação:

$\mathbf{Y}_{i(T \times M_i)}$ —matriz das observações das  $M_i$  variáveis endógenas que aparecem na equação  $i$ ,

$\mathbf{Y}_{i(T \times M_i^*)}^*$ —matriz das observações das  $M_i^*$  variáveis endógenas que não aparecem na equação  $i$ ,

$\mathbf{X}_{i(T \times k_i)}$ —matriz das observações das  $k_i$  variáveis exógenas que aparecem na equação  $i$ ,

$\mathbf{X}_{i(T \times k_i^*)}^*$ —matriz das observações das  $k_i^*$  variáveis exógenas que não aparecem na equação  $i$ ,

$\gamma_{i(M_i \times 1)}$ —vector dos coeficientes das variáveis endógenas que aparecem na equação  $i$ , exceptuando a variável que foi normalizada  $\mathbf{y}_i$ ,

$\gamma_{i(M_i^* \times 1)}^*$ —vector dos coeficientes das variáveis endógenas que não aparecem na equação  $i$ ; é um vector nulo,

$\beta_{i(k_i \times 1)}$ —vector dos coeficientes das variáveis exógenas que aparecem na equação  $i$ ,



	V.endógenas	Nº de variáveis	V.exógenas	Nº de variáveis
Incluídas na equação $i$	$\mathbf{Y}_i$	$M_i$	$\mathbf{X}_i$	$k_i$
Excluídas da equação $i$	$\mathbf{Y}_i^*$	$M_i^*$	$\mathbf{X}_i^*$	$k_i^*$
Total		$M_i + M_i^* = M$		$k_i + k_i^* = k$

Tabela 2.1: Partição relativa às variáveis que aparecem na equação  $i$ .

$\beta_{i(k_i^* \times 1)}^*$ —vector dos coeficientes das variáveis exógenas que não aparecem na equação  $i$ ; é um vector nulo,

$\Gamma_i$  e  $\mathbf{B}_i$  —  $i$ -ésimas colunas das matrizes,  $\mathbf{\Gamma}$  e  $\mathbf{B}$ , respectivamente.

Deste modo, em termos matriciais, pode agora considerar-se que  $\mathbf{Y} = \begin{bmatrix} \mathbf{y}_i & \mathbf{Y}_i & \mathbf{Y}_i^* \end{bmatrix}$ ,  $\mathbf{X} = \begin{bmatrix} \mathbf{X}_i & \mathbf{X}_i^* \end{bmatrix}$ ; relativamente às matrizes de parâmetros, vem a partição

$$\Gamma_i = \begin{bmatrix} -1 \\ \gamma_i \\ \gamma_i^* \end{bmatrix} = \begin{bmatrix} -1 \\ \gamma_i \\ \mathbf{0} \end{bmatrix},$$

$$\mathbf{B}_i = \begin{bmatrix} \beta_i \\ \beta_i^* \end{bmatrix} = \begin{bmatrix} \beta_i \\ \mathbf{0} \end{bmatrix}.$$

Esta partição vai ter consequências na escrita da equação  $i$ , a qual pode ser expressa apenas em termos dos blocos de coeficientes não nulos, lembrando que a equação  $i$  da forma estrutural se escreve  $\mathbf{Y}\Gamma_i + \mathbf{X}\mathbf{B}_i + \mathbf{e}_i = \mathbf{0}$ , onde  $\Gamma_i$  e  $\mathbf{B}_i$  representam as colunas  $i$  das matrizes  $\mathbf{\Gamma}$  e  $\mathbf{B}$ , respectivamente (incluindo os elementos nulos). Substituindo as matrizes de parâmetros de acordo com a partição anterior, tem-se que

$$\mathbf{y}_i = \mathbf{Y}_i\gamma_i + \mathbf{Y}_i^*\gamma_i^* + \mathbf{X}_i\beta_i + \mathbf{X}_i^*\beta_i^* + \mathbf{e}_i \quad (2.3.2)$$

$$\Leftrightarrow \mathbf{y}_i = \mathbf{Y}_i\gamma_i + \mathbf{X}_i\beta_i + \mathbf{e}_i. \quad (2.3.3)$$

Na Tabela 2.1, resume-se a notação relativa aos blocos correspondentes à equação  $i$  da forma estrutural.

Do mesmo modo, as restrições de exclusão induzem uma decomposição por blocos nas matrizes da forma reduzida.

Definindo como  $\mathbf{Z}_i = \begin{bmatrix} \mathbf{Y}_i & \mathbf{X}_i \end{bmatrix}_{(T \times (M_i + k_i))}$  a matriz por blocos com as observações das variáveis endógenas e exógenas que aparecem na equação  $i$  e por

$\boldsymbol{\delta}_i = \begin{bmatrix} \boldsymbol{\gamma}'_i & \boldsymbol{\beta}'_i \end{bmatrix}'_{((M_i+k_i) \times 1)}$  o vector dos  $(k_i + M_i)$  parâmetros estruturais da equação  $i$ , obtém-se a seguinte forma para a equação  $i$ ,

$$\mathbf{y}_i = \mathbf{Z}_i \boldsymbol{\delta}_i + \mathbf{e}_i, i = 1, \dots, M. \quad (2.3.4)$$

Em relação à matriz de parâmetros da forma reduzida,  $\boldsymbol{\Pi} = -\mathbf{B}\boldsymbol{\Gamma}^{-1}$ , note-se que ela pode ser escrita sob a forma de um produto de matrizes definidas por blocos, do seguinte modo:

$$\underbrace{\begin{bmatrix} \boldsymbol{\Pi} & \mathbf{I}_k \end{bmatrix}}_{(k \times (M+k))} \underbrace{\begin{bmatrix} \boldsymbol{\Gamma} \\ \mathbf{B} \end{bmatrix}}_{((M+k) \times M)} = \underbrace{\mathbf{0}}_{(k \times M)}.$$

Uma vez que  $\boldsymbol{\Gamma}_i$  e  $\mathbf{B}_i$  correspondem à coluna  $i$  das matrizes  $\boldsymbol{\Gamma}$  e  $\mathbf{B}$ , respectivamente, tem-se para a equação  $i$  que

$$\begin{bmatrix} \boldsymbol{\Pi} & \mathbf{I}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{\Gamma}_i \\ \mathbf{B}_i \end{bmatrix} = \mathbf{0}. \quad (2.3.5)$$

Como a matriz  $\begin{bmatrix} \boldsymbol{\Pi} & \mathbf{I}_k \end{bmatrix}$  tem característica  $k$ , a equação matricial (2.3.5) corresponde a um sistema de  $k$  equações e  $(M + k)$  incógnitas. Havendo maior número de incógnitas do que de equações, torna-se necessário dispor de informação adicional, mais especificamente, de  $M$  equações adicionais, para possibilitar a identificação dos parâmetros, (ou seja, das componentes de  $\boldsymbol{\Gamma}_i$  e de  $\mathbf{B}_i$ ), a partir da informação contida em  $\boldsymbol{\Pi}$ . Para esse efeito, podem ser usadas as equações lineares que traduzem as restrições de exclusão e a regra de normalização.

Estão agora reunidas as condições para apresentar alguns critérios que podem ser usados para verificar se uma equação de um sistema é identificada.

Para isso, note-se que as restrições de exclusão para a equação  $i$  se podem escrever como

$$\mathbf{R}_i \begin{bmatrix} \boldsymbol{\Gamma}_i \\ \mathbf{B}_i \end{bmatrix} = \mathbf{R}_i \boldsymbol{\Delta}_i = \mathbf{0}, \quad (2.3.6)$$

onde  $\boldsymbol{\Delta}_i = \begin{bmatrix} \boldsymbol{\Gamma}_i \\ \mathbf{B}_i \end{bmatrix}$  e  $\mathbf{R}_i$  é uma matriz conhecida, com característica  $J$ , com  $J < M + k$ , sendo  $J$  o número de restrições de exclusão da  $i$ -ésima equação. Repare-se que cada linha de  $\mathbf{R}_i$  (que tem apenas um elemento não nulo, igual à unidade)

representa uma restrição de exclusão, pois isso implica que um determinado parâmetro seja nulo, ou seja, que a variável correspondente não apareça na equação  $i$ . Note-se ainda que se  $J$  representar o número de restrições de exclusão da  $i$ -ésima equação, então, de acordo com (2.3.2) e com a Tabela 2.1, vem

$$J = M_i^* + k_i^*.$$

Por outro lado,  $J$  é a característica de  $\mathbf{R}_{i(J \times (M+k))}$ , logo,

$$\text{rank}(\mathbf{R}_i) = M_i^* + k_i^*, \quad (2.3.7)$$

onde  $\text{rank}(\mathbf{R}_i)$  designa a característica da matriz  $\mathbf{R}_i$ .

Os critérios que facilitam a verificação da identificabilidade para a equação  $i$  baseiam-se em (2.3.5) e (2.3.6). Tendo em conta que (2.3.5) contém  $k$  equações sobre as  $M + k$  incógnitas, no caso de (2.3.6) contribuir com  $M - 1$  equações e juntando a equação que estabelece a regra de normalização da equação  $i$ , reúnem-se  $M + k$  equações, ou seja, há informação suficiente para que os  $M + k$  parâmetros estruturais da equação  $i$  possam ser expressos a partir dos parâmetros da forma reduzida. Deste modo, assegura-se que os parâmetros da equação  $i$  são identificados.

Conjugando então a informação de (2.3.5) e (2.3.6) com o facto de  $\mathbf{R}_i$  ter característica  $M_i^* + k_i^*$ , vem que

$$\underbrace{\left[ \begin{array}{c} \left( \begin{array}{ccc} \mathbf{\Pi} & \vdots & \mathbf{I}_k \end{array} \right) \\ \mathbf{R}_i \end{array} \right]}_{((k+J) \times (M+k))} \underbrace{\Delta_i}_{((M+k) \times 1)} = \underbrace{\mathbf{0}}_{((k+J) \times 1)}. \quad (2.3.8)$$

Mas, usando resultados gerais sobre sistemas de equações lineares, conclui-se que a característica de

$$\left[ \begin{array}{c} \left( \begin{array}{ccc} \mathbf{\Pi} & \vdots & \mathbf{I}_k \end{array} \right) \\ \mathbf{R}_i \end{array} \right] \quad (2.3.9)$$

deve ser  $(k + M - 1)$ , de forma a garantir que as equações de (2.3.5) são independentes das de (2.3.6). Consegue-se, assim, identificar  $\Delta_i$  em termos de  $\mathbf{\Pi}$ .

Em termos práticos, para determinar a característica da matriz (2.3.9), pode recorrer-se às seguintes proposições, cujas demonstrações podem ser vistas em Schmidt (1976).

**Proposição 2.3.1.** Considerando  $\Delta = \begin{bmatrix} \Gamma \\ \mathbf{B} \end{bmatrix}$ , a característica de (2.3.9) é  $(k + M - 1)$  se e só se a característica de  $(\mathbf{R}_i \Delta)$  for  $M - 1$ .  $\square$

Surge assim uma seguinte condição necessária e suficiente para a identificabilidade da equação  $i$ .

**Proposição 2.3.2. (Condição de característica)** É condição necessária e suficiente para a equação  $i$  ser identificada, usando as restrições  $\mathbf{R}_i \Delta_i = \mathbf{0}$ , que

$$\text{rank}(\mathbf{R}_i \Delta) = M - 1,$$

com  $\Delta = \begin{bmatrix} \Gamma' & \mathbf{B}' \end{bmatrix}'$  e  $\mathbf{R}_i$  definido em (2.3.6).  $\square$

Tendo em conta que  $\mathbf{R}_i \Delta$  tem  $M$  colunas e que a  $i$ -ésima coluna,  $\mathbf{R}_i \Delta_i$ , só tem elementos nulos, para que  $\mathbf{R}_i \Delta$  tenha característica  $M - 1$ , deve ter pelo menos  $M - 1$  linhas, ou seja, a característica de  $\mathbf{R}_i$  deve ser no mínimo  $M - 1$ . Isto pode traduzir-se pela seguinte condição, que é uma condição necessária para a identificabilidade da equação  $i$ .

**Proposição 2.3.3. (Condição de ordem)** Se a equação  $i$  for identificada então

$$\text{rank}(\mathbf{R}_i) \geq M - 1. \quad (2.3.10)$$

$\square$

Conjugando (2.3.10) com o facto de a característica de  $\mathbf{R}_i$  verificar (2.3.7), resulta que quando uma equação  $i$  é identificada se verifica que

$$M_i^* + k_i^* \geq M - 1 \Leftrightarrow k_i^* \geq M_i - 1. \quad (2.3.11)$$

A proposição seguinte resume as diversas situações que podem ocorrer no estudo da identificabilidade de uma equação.

**Proposição 2.3.4.** O estudo da identificabilidade de uma equação  $i$ , conduz a uma das situações:

1. A equação  $i$  não é identificada, dizendo-se que é **subidentificada**, se

$$\text{rank}(\mathbf{R}_i) < M - 1.$$

Neste caso não é possível obter os parâmetros estruturais da equação  $i$  a partir dos parâmetros da forma reduzida.

2. A equação  $i$  é **exactamente identificada** se

$$\text{rank}(\mathbf{R}_i\mathbf{\Delta}) = M - 1 \text{ e } \text{rank}(\mathbf{R}_i) = M - 1.$$

Neste caso os parâmetros estruturais da equação  $i$  podem ser obtidos a partir dos parâmetros da forma reduzida de uma forma única.

3. A equação  $i$  é **sobreidentificada** se

$$\text{rank}(\mathbf{R}_i\mathbf{\Delta}) = M - 1 \text{ e } \text{rank}(\mathbf{R}_i) > M - 1.$$

Neste caso existem diferentes formas de exprimir os parâmetros estruturais da equação  $i$  em função dos parâmetros da forma reduzida.  $\square$

Para ilustrar a aplicação dos critérios referidos, apresenta-se um outro exemplo publicado em Judge *et al.* (1988).

*Exemplo 2.3.1.* Considere-se um *SEM* definido por um sistema de quatro equações ( $M = 4$ ) envolvendo quatro variáveis endógenas e quatro variáveis exógenas ( $k = 4$ )

$$\left\{ \begin{array}{llllll} \nu_{11}\mathbf{y}_1 + \nu_{21}\mathbf{y}_2 & + \nu_{41}\mathbf{y}_4 + \beta_{11}\mathbf{x}_1 & + \beta_{41}\mathbf{x}_4 + \mathbf{e}_1 & = & \mathbf{0} \\ \nu_{12}\mathbf{y}_1 + \nu_{22}\mathbf{y}_2 & + \beta_{12}\mathbf{x}_1 + \beta_{22}\mathbf{x}_2 & + \mathbf{e}_2 & = & \mathbf{0} \\ \nu_{13}\mathbf{y}_1 + \nu_{23}\mathbf{y}_2 + \nu_{33}\mathbf{y}_3 + \nu_{43}\mathbf{y}_4 + \beta_{13}\mathbf{x}_1 & + \beta_{43}\mathbf{x}_4 + \mathbf{e}_3 & = & \mathbf{0} \\ & \nu_{44}\mathbf{y}_4 + \beta_{14}\mathbf{x}_1 + \beta_{24}\mathbf{x}_2 + \beta_{34}\mathbf{x}_3 + \beta_{44}\mathbf{x}_4 + \mathbf{e}_4 & = & \mathbf{0} \end{array} \right.$$

Passando à forma matricial estrutural,  $\mathbf{Y}\mathbf{\Gamma} + \mathbf{X}\mathbf{B} + \mathbf{E} = \mathbf{0}$ , tem-se que

$$\mathbf{\Gamma} = \begin{bmatrix} \nu_{11} & \nu_{12} & \nu_{13} & 0 \\ \nu_{21} & \nu_{22} & \nu_{23} & 0 \\ 0 & 0 & \nu_{33} & 0 \\ \nu_{41} & 0 & \nu_{43} & \nu_{44} \end{bmatrix}, \mathbf{B} = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} & \beta_{14} \\ 0 & \beta_{22} & 0 & \beta_{24} \\ 0 & 0 & 0 & \beta_{34} \\ \beta_{41} & 0 & \beta_{43} & \beta_{44} \end{bmatrix}.$$

Para estudar a identificabilidade, considera-se uma equação de cada vez, calcula-se a correspondente matriz  $\mathbf{R}_i$ , formada pelos coeficientes das restrições de exclusão nessa equação; de seguida classifica-se a equação, tendo em conta que  $\mathbf{\Delta} = \begin{bmatrix} \mathbf{\Gamma} \\ \mathbf{B} \end{bmatrix}$  e aplicando

a **Proposição 2.3.4.**

Assim, para a primeira equação, tem-se que as variáveis que têm coeficiente nulo são  $y_3, x_2$  e  $x_3$ , logo

$$\mathbf{R}_1 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, e \mathbf{R}_1 \Delta = \begin{bmatrix} 0 & 0 & \nu_{33} & 0 \\ 0 & \beta_{22} & 0 & \beta_{24} \\ 0 & 0 & 0 & \beta_{34} \end{bmatrix},$$

pelo que  $rank(\mathbf{R}_1 \Delta) = 3 = (M - 1) = rank(\mathbf{R}_1)$ , logo a primeira equação é exactamente identificada.

Para a segunda equação, as variáveis excluídas são  $y_3, y_4, x_3$  e  $x_4$ . Deste modo, tem-se que

$$\mathbf{R}_2 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} e \text{ que } \mathbf{R}_2 \Delta = \begin{bmatrix} 0 & 0 & \nu_{33} & 0 \\ \nu_{41} & 0 & \nu_{43} & \nu_{44} \\ 0 & 0 & 0 & \beta_{34} \\ \beta_{41} & 0 & \beta_{43} & \beta_{44} \end{bmatrix};$$

pelo que se conclui que  $rank(\mathbf{R}_2 \Delta) = 3$  e  $rank(\mathbf{R}_2) = 4 > M - 1$ . Assim, a segunda equação é sobreidentificada.

Quanto à terceira equação, as variáveis não incluídas são  $x_2$  e  $x_3$ ; logo,

$$\mathbf{R}_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix},$$

cuja característica é  $2 < M - 1$ , pelo que esta equação é subidentificada.

Finalmente para a quarta equação, cujas variáveis excluídas são  $y_1, y_2$  e  $y_3$ , tem-se que

$$\mathbf{R}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} e \text{ que } \mathbf{R}_4 \Delta = \begin{bmatrix} \nu_{11} & \nu_{12} & \nu_{13} & 0 \\ \nu_{21} & \nu_{22} & \nu_{23} & 0 \\ 0 & 0 & \nu_{33} & 0 \end{bmatrix}.$$

Deste modo, esta equação é exactamente identificada, uma vez que  $rank(\mathbf{R}_4 \Delta) = 3 = (M - 1) = rank(\mathbf{R}_4)$ .  $\diamond$

## Capítulo 3

# Estimação tradicional dos parâmetros de um *SEM*

O presente capítulo sumaria os processos tradicionalmente usados para estimar os parâmetros dos *SEM*. O método mais vulgarizado baseia-se no princípio dos mínimos quadrados. Por esse motivo se apresentam as diferentes versões de estimadores de mínimos quadrados existentes para estimar um *SEM*, nomeadamente, mínimos quadrados ordinários (*OLS* de *ordinary least squares*), mínimos quadrados indirectos (*ILS* de *indirect least squares*), mínimos quadrados em duas fases (*2SLS* de *two-stage least squares*) e mínimos quadrados em três fases (*3SLS* de *three-stage least squares*). Estas versões permitiram melhorar sucessivamente as propriedades do estimador de mínimos quadrados na sua forma mais simples.

De seguida, estuda-se o Método dos Momentos Generalizado (*GMM* de *Generalized Method of Moments*). Este último método é bem mais recente, mas é considerado vantajoso para o tipo de modelos em questão e já está integrado nos principais livros de texto introdutório de Econometria; terá uma importância fundamental no desenvolvimento dos capítulos seguintes, de acordo com os objectivos da presente dissertação.

O Método da Máxima Verosimilhança, que é também tratado por alguns autores, exige o conhecimento das distribuições de todas as variáveis envolvidas no sistema e acaba por ter menos interesse no estudo dos *SEM*. Assim, não será considerado no presente trabalho.

Ao estimar os parâmetros do modelo seria natural considerar conjuntamente todos os parâmetros incluídos no sistema. No entanto, historicamente não foi isso que aconteceu. De facto, por qualquer dos métodos (mínimos quadrados ou momentos generalizados), a estimação dos parâmetros começou por ser feita equação a equação (informação limitada). Posteriormente foram desenvolvidos estimadores obtidos conjuntamente para todos os parâmetros do sistema (informação completa). Daí resulta que, qualquer que seja o princípio que preside à estimação, existam métodos de informação limitada e de informação completa.

Neste capítulo apresentam-se os estimadores obtidos pelos diferentes métodos e as respectivas propriedades.

### 3.1 Informação limitada e informação completa

No capítulo anterior foi já destacada a importância da identificabilidade. Lembra-se que se uma equação não for identificada, isso significa que não há informação suficiente para determinar algebricamente os parâmetros dessa equação e que, portanto, não faz sentido estimar esses parâmetros. Assim, para passar ao estudo dos estimadores, admite-se que todas as equações do modelo são identificadas, podendo ser exactamente identificadas ou sobreidentificadas.

Tendo em conta que um *SEM* é definido por um sistema de equações, a estimação dos parâmetros pode fazer-se separadamente para cada equação ou conjuntamente para todas as equações do sistema. Deste modo, os métodos de estimação usuais nestes modelos podem dividir-se em dois grupos, os Métodos de Informação Limitada (*MIL*) - em que os parâmetros de cada equação são estimados separadamente, desprezando a informação contida nas restantes equações - e os Métodos de Informação Completa (*MIC*) - em que os parâmetros de todas as equações do modelo são estimados simultaneamente.

Existe alguma controvérsia em relação à opção por *MIL* ou por *MIC*. Como os *MIC* estimam conjuntamente todos os parâmetros, a correlação existente entre os erros das diferentes equações está presente no processo de estimação e os estimadores gozam de boas propriedades assintóticas, sendo geralmente considerados superiores aos *MIL*. No entanto, os *MIC* apresentam também algumas desvantagens. De facto, do ponto de



	<i>MIL</i>	<i>MIC</i>
Método dos Mínimos Quadrados	<i>2SLS</i>	<i>3SLS</i>
Método dos Momentos Generalizado	<i>GMMIL</i>	<i>GMMIC</i>
Método da Máxima Verosimilhança	<i>MLIL</i>	<i>MLIC</i>

Tabela 3.1: Estimadores usuais em *SEM*.

vista prático, os *MIC* envolvem cálculos muito mais pesados; para além disso, uma falha de modelação numa equação vai afectar a estimação dos parâmetros de todas as equações.

Assim, as duas metodologias, *MIL* e *MIC*, são apresentadas na literatura, mas não se torna clara qual é a opção mais vantajosa em termos práticos. Por exemplo, Greene (2003) considera que, em amostras de dimensão moderada, nem sempre se confirma a superioridade dos *MIC*. Daí que seja da maior importância a realização de um estudo comparativo das propriedades dos estimadores obtidos com *MIL* ou com *MIC*. Esse estudo está publicado em Judge *et al.* (1988) para estimadores de mínimos quadrados, comparando o estimador *2SLS* com o estimador *3SLS*, mas não inclui os estimadores *GMM*. Neste trabalho efectuou-se a comparação correspondente para os estimadores *GMM*, a qual foi possível a partir da obtenção de formas apropriadas das matrizes de covariâncias assintóticas dos estimadores. Os resultados serão apresentados na **Secção 3.9**.

Dada a grande diversidade de métodos de estimação e respectivos estimadores, eles serão identificados pelas siglas constantes na Tabela 3.1.

## 3.2 Utilização do Método dos Mínimos Quadrados em *SEM*

Atendendo à semelhança já referida existente entre os *SEM* e os *MRLM*, o método dos mínimos quadrados surge como um processo natural para estimar os parâmetros de um *SEM*. No entanto, importa ter em atenção que um *SEM* não verifica algumas das hipóteses assumidas para o *MRLM*.

A estimação dos parâmetros da forma reduzida pode fazer-se aplicando o *OLS*, uma

vez que nessa forma todos os regressores são variáveis exógenas e que estas são não correlacionadas com os erros. Mas, os parâmetros com maior interesse na aplicação de um *SEM* são os estruturais. No entanto, os estimadores de mínimos quadrados dos parâmetros estruturais, obtidos aplicando *OLS* à forma estrutural, não são consistentes, pois nessa forma há variáveis endógenas no conjunto dos regressores, as quais são correlacionadas com os erros aleatórios. Assim, uma solução para ultrapassar este problema consiste em procurar estimadores consistentes para os parâmetros da forma reduzida, por *OLS* e, a partir deles, determinar estimadores para os parâmetros estruturais, com base nas relações (2.2.12) e (2.2.13). Note-se que este procedimento nem sempre é possível, ficando dependente das condições de identificabilidade das equações. Quando é possível, conduz a um estimador que é designado por estimador de mínimos quadrados indirectos (*ILS*) e que será estudado numa secção a seguir.

Chama-se ainda à atenção para o facto de ao proceder à estimação dos parâmetros de um *SEM* por *OLS*, ser necessário assumir que os erros de cada equação são homocedásticos em relação ao tempo, ou seja, assumir que

$$E[e_{it}e_{it}] = \sigma_{it}^2 = \sigma_i^2 \text{ e que } E[e_{it}e_{jt}] = \sigma_{ijt} = \sigma_{ij}, \forall t = 1, \dots, T.$$

Neste caso, de acordo com (2.2.6) e (2.2.7), tem-se que

$$E[\mathbf{e}_i\mathbf{e}_i'] = \sigma_i^2\mathbf{I}_T = \boldsymbol{\Omega}_i \text{ e } E[\mathbf{e}_i\mathbf{e}_j'] = \sigma_{ij}\mathbf{I}_T = \boldsymbol{\Omega}_{ij}. \quad (3.2.1)$$

Deste modo, considerando  $\mathbf{e} = \text{vec}(\mathbf{E})$ , é possível escrever a matriz de covariâncias dos erros da forma estrutural,  $\boldsymbol{\Omega} = [\boldsymbol{\Omega}_{ij}]_{i,j=1,\dots,M} = V[\mathbf{e}]$ , como um produto de Kronecker, uma vez que

$$\begin{aligned} \boldsymbol{\Omega} &= V[\mathbf{e}] = E[\mathbf{e}\mathbf{e}'] \\ &= E \begin{bmatrix} \mathbf{e}_1\mathbf{e}_1' & \mathbf{e}_1\mathbf{e}_2' & \cdots & \mathbf{e}_1\mathbf{e}_M' \\ \mathbf{e}_2\mathbf{e}_1' & \mathbf{e}_2\mathbf{e}_2' & \cdots & \mathbf{e}_2\mathbf{e}_M' \\ \vdots & \vdots & & \vdots \\ \mathbf{e}_M\mathbf{e}_1' & \mathbf{e}_M\mathbf{e}_2' & \cdots & \mathbf{e}_M\mathbf{e}_M' \end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} \Omega_{11} & \Omega_{12} & \cdots & \Omega_{1M} \\ \Omega_{21} & \Omega_{22} & \cdots & \Omega_{2M} \\ \vdots & \vdots & & \vdots \\ \Omega_{M1} & \Omega_{M2} & \cdots & \Omega_{MM} \end{bmatrix} \quad (3.2.2)$$

$$= \begin{bmatrix} \sigma_1^2 \mathbf{I}_T & \sigma_{12} \mathbf{I}_T & \cdots & \sigma_{1M} \mathbf{I}_T \\ \sigma_{21} \mathbf{I}_T & \sigma_2^2 \mathbf{I}_T & \cdots & \sigma_{2M} \mathbf{I}_T \\ \vdots & \vdots & & \vdots \\ \sigma_{M1} \mathbf{I}_T & \sigma_{M2} \mathbf{I}_T & \cdots & \sigma_M^2 \mathbf{I}_T \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1M} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2M} \\ \vdots & \vdots & & \vdots \\ \sigma_{M1} & \sigma_{M2} & \cdots & \sigma_M^2 \end{bmatrix} \otimes \mathbf{I}_T$$

$$= (\Sigma \otimes \mathbf{I}_T) = \Omega. \quad (3.2.3)$$

A adaptação do princípio dos mínimos quadrados a um *SEM* no caso geral (sem impor homocedasticidade) será tratada em secções seguintes.

### 3.3 Estimadores de mínimos quadrados com informação limitada

Conforme se referiu anteriormente, de acordo com os métodos de informação limitada, procede-se à estimação dos parâmetros de cada uma das equações, sucessivamente. Assim, não é tomada em consideração a informação contida nas variáveis endógenas que aparecem apenas nas restantes equações do sistema, nem a possível interacção existente entre as equações. É por esta razão que se diz que estes métodos são de informação limitada.

Ao estimar separadamente as equações, desprezam-se as correlações existentes entre os erros das diferentes equações, ou seja, considera-se que  $\sigma_{ij} = 0, i \neq j$ . Deste modo, os blocos da matriz  $\Omega$  na forma (3.2.2) que não pertencem à diagonal principal são nulos. Assim, verifica-se a expressão (3.2.3) com  $\Sigma = \text{diag} \left[ \sigma_1^2 \quad \sigma_2^2 \quad \cdots \quad \sigma_M^2 \right]$ , ou

seja, conclui-se que

$$\begin{aligned}\boldsymbol{\Omega} &= \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_M^2 \end{bmatrix} \otimes \mathbf{I}_T \\ &= \text{diag} \left[ \sigma_1^2 \quad \sigma_2^2 \quad \cdots \quad \sigma_M^2 \right] \otimes \mathbf{I}_T.\end{aligned}\tag{3.3.1}$$

### 3.3.1 Estimador *OLS*

Considere-se um *SEM* na forma reduzida, ou seja,  $\mathbf{Y} = \mathbf{X}\boldsymbol{\Pi} + \mathbf{V}$ , definida em (2.2.11).

Destacando a equação  $i$ , verifica-se que

$$\mathbf{y}_i = \mathbf{X}\boldsymbol{\pi}_i + \mathbf{v}_i, i = 1, \dots, M.\tag{3.3.2}$$

O relacionamento entre  $\mathbf{X}$  e o termo aleatório dos erros  $\mathbf{V}$  é semelhante ao que existe no *MRLM*. O estimador de mínimos quadrados para o vector de parâmetros  $\boldsymbol{\pi}_i$  da forma reduzida vem directamente

$$\hat{\boldsymbol{\pi}}_i = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}_i.$$

Notando que (3.3.2) permite escrever  $\hat{\boldsymbol{\pi}}_i$  noutra forma,

$$\begin{aligned}\hat{\boldsymbol{\pi}}_i &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}_i \\ &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{X}\boldsymbol{\pi}_i + \mathbf{v}_i) \\ &= \boldsymbol{\pi}_i + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{v}_i,\end{aligned}\tag{3.3.3}$$

facilmente se verifica que o estimador é centrado e consistente para  $\boldsymbol{\pi}_i$ .

Apesar de geralmente ser dada mais importância à estimação dos parâmetros estruturais do que à estimação dos parâmetros da forma reduzida, a segunda situação também tem interesse, na medida em que permite avaliar o impacto médio provocado na variável endógena (considerada dependente em cada equação da forma reduzida), por um acréscimo unitário verificado em cada variável exógena que apareça nessa equação. Para compreender melhor esta ideia, considere-se, por exemplo, a equação  $y_t = \pi_0 + \pi_1 x_t + v_t$ , retirada de um *SEM* na forma reduzida, onde a variável endógena,  $y_t$ , representa o rendimento e a variável exógena,  $x_t$ , corresponde ao investimento no

período  $t$ . Se a estimativa de  $\pi_1$  por *OLS* fosse  $\hat{\pi}_1 = 5$ , poderia dizer-se que, em termos médios, um crescimento de uma unidade no investimento, seria acompanhado por um aumento de cinco unidades no rendimento, no mesmo período de tempo.

Para além deste tipo de interpretações, há interesse na estimação da forma reduzida, por *OLS*, quando se pretende efectuar predições sobre as variáveis endógenas.

Considere-se agora a estimação dos parâmetros da forma estrutural.

Partindo da equação  $i$  na forma estrutural obtida em (2.3.4), ou seja,  $\mathbf{y}_i = \mathbf{Z}_i \boldsymbol{\delta}_i + \mathbf{e}_i$ , pretende-se estimar  $\boldsymbol{\delta}_i$  aplicando o método dos mínimos quadrados. Como se mostrará, se for ignorada a correlação entre  $\mathbf{Z}_i$  e  $\mathbf{e}_i$ , a aplicação de *OLS* conduz a um estimador enviesado e inconsistente.

*Exemplo 3.3.1.* Para ilustrar a falta de consistência do estimador *OLS* e seguindo Gujarati (2003), considere-se de novo o modelo de Keynes apresentado em (2.1.1) e suponha-se que se pretende estimar o parâmetro  $\gamma$  da segunda equação

$$c_t = \beta + \gamma y_t + \varepsilon_t. \quad (3.3.4)$$

Admitindo que o termo de erro aleatório é tal que,  $E[\varepsilon_t] = 0$ , que  $E[\varepsilon_t^2] = \sigma^2$ , que  $E[\varepsilon_t \varepsilon_{t+j}] = 0, j \neq 0$  e que  $Cov(x_t, \varepsilon_t) = 0$ , vai mostrar-se que  $y_t$  e  $\varepsilon_t$  são correlacionados; posteriormente, mostrar-se-á que o estimador  $\hat{\gamma}_{OLS}$  é inconsistente para estimar  $\gamma$ . Ao substituir a expressão da segunda equação na primeira, resulta

$$\begin{aligned} y_t &= \beta + \gamma y_t + x_t + \varepsilon_t \\ \Leftrightarrow y_t &= \frac{\beta}{1-\gamma} + \frac{1}{1-\gamma} x_t + \frac{1}{1-\gamma} \varepsilon_t. \end{aligned}$$

Como  $E[\varepsilon_t] = 0$  e  $x_t$  é exógena, vem que  $E[y_t] = \frac{\beta}{1-\gamma} + \frac{1}{1-\gamma} x_t$ . Assim, resulta que  $y_t - E[y_t] = \frac{1}{1-\gamma} \varepsilon_t$ . Mas, por outro lado,  $\varepsilon_t - E[\varepsilon_t] = \varepsilon_t$ , pelo que

$$Cov(y_t, \varepsilon_t) = E[(y_t - E[y_t])(\varepsilon_t - E[\varepsilon_t])] = \frac{E[\varepsilon_t^2]}{1-\gamma} = \frac{\sigma^2}{1-\gamma} > 0, \quad (3.3.5)$$

de acordo com a hipótese. Conclui-se assim que as variáveis  $y_t$  e  $\varepsilon_t$  são correlacionadas.

Para mostrar que o estimador  $\hat{\gamma}_{OLS}$  não é consistente para  $\gamma$ , interessa recordar que o estimador *OLS* para  $\gamma$  é da forma  $\hat{\gamma}_{OLS} = \frac{\sum_t C_t Y_t}{\sum_t Y_t^2}$ , com  $C_t = (c_t - \bar{c})$  e  $Y_t = (y_t - \bar{y})$ .

Deste modo, substituindo  $c_t$  de acordo com (3.3.4) e procedendo a alguns cálculos algébricos, vem que

$$\hat{\gamma}_{OLS} = \frac{\sum_t (\beta + \gamma y_t + \varepsilon_t) Y_t}{\sum_t Y_t^2} = \gamma + \frac{\sum_t \varepsilon_t Y_t}{\sum_t Y_t^2}.$$

Calculando agora o limite em probabilidade de  $\hat{\gamma}_{OLS}$  e tendo em conta que a variância e a covariância amostrais convergem, respectivamente, para a variância e a covariância populacionais e que  $\bar{\varepsilon} = 0$ , resulta que

$$\begin{aligned} p \lim_{T \rightarrow \infty} \hat{\gamma}_{OLS} &= \gamma + p \lim_{T \rightarrow \infty} \left[ \frac{\sum_t \varepsilon_t Y_t}{\sum_t Y_t^2} \right] \\ &= \gamma + p \lim_{T \rightarrow \infty} \left[ \frac{\frac{1}{T} \sum_t \varepsilon_t (y_t - \bar{y})}{\frac{1}{T} \sum_t Y_t^2} \right] \\ &= \gamma + p \lim_{T \rightarrow \infty} \left[ \frac{\frac{1}{T} \sum_t \varepsilon_t y_t}{\frac{1}{T} \sum_t Y_t^2} \right] \\ &= \gamma + \frac{\left( \frac{\sigma^2}{1-\gamma} \right)}{\sigma_y^2}. \end{aligned}$$

Como a segunda parcela é positiva, de acordo com (3.3.5), conclui-se que  $p \lim_{T \rightarrow \infty} \hat{\gamma}_{OLS} \neq \gamma$ , o que traduz que  $\hat{\gamma}_{OLS}$  não é um estimador consistente para  $\gamma$ .  $\diamond$

Ao estimar os parâmetros de uma equação da forma estrutural, aplicando directamente o método dos mínimos quadrados à equação (2.3.4), obtêm-se estimadores que, no caso geral, não são centrados, nem consistentes, como se verá a seguir.

Ignorando a correlação entre  $\mathbf{Z}_i$  e  $\mathbf{e}_i$ , o vector de parâmetros  $\boldsymbol{\delta}_i$  seria estimado por

$$\hat{\boldsymbol{\delta}}_{i,OLS} = (\mathbf{Z}_i' \mathbf{Z}_i)^{-1} \mathbf{Z}_i' \mathbf{y}_i.$$

Para verificar que este estimador é enviesado para  $\boldsymbol{\delta}_i$ , calcule-se o valor esperado

de  $\hat{\boldsymbol{\delta}}_{i,OLS}$ .

$$\begin{aligned}
E \left[ \hat{\boldsymbol{\delta}}_{i,OLS} \right] &= E \left[ (\mathbf{Z}'_i \mathbf{Z}_i)^{-1} \mathbf{Z}'_i \mathbf{y}_i \right] \\
&= E \left[ (\mathbf{Z}'_i \mathbf{Z}_i)^{-1} \mathbf{Z}'_i (\mathbf{Z}_i \boldsymbol{\delta}_i + \mathbf{e}_i) \right] \\
&= E \left[ \underbrace{(\mathbf{Z}'_i \mathbf{Z}_i)^{-1} \mathbf{Z}'_i \mathbf{Z}_i}_{\mathbf{I}} \boldsymbol{\delta}_i \right] + E \left[ (\mathbf{Z}'_i \mathbf{Z}_i)^{-1} \mathbf{Z}'_i \mathbf{e}_i \right] \\
&= \boldsymbol{\delta}_i + E \left[ (\mathbf{Z}'_i \mathbf{Z}_i)^{-1} \mathbf{Z}'_i \mathbf{e}_i \right].
\end{aligned} \tag{3.3.6}$$

Como  $\mathbf{Z}_i = \begin{bmatrix} \mathbf{Y}_i & \mathbf{X}_i \end{bmatrix}$  contém os valores de variáveis endógenas,  $\mathbf{Y}_i$ , definidas pela partição apresentada na Tabela 2.1, que são conjuntamente determinados com  $\mathbf{y}_i$ , pode concluir-se que  $\mathbf{Z}_i$  e os erros aleatórios  $\mathbf{e}_i$  são correlacionados. Deste modo,

$$E \left[ (\mathbf{Z}'_i \mathbf{Z}_i)^{-1} \mathbf{Z}'_i \mathbf{e}_i \right] \neq \mathbf{0} \Rightarrow E \left[ \hat{\boldsymbol{\delta}}_{i,OLS} \right] \neq \boldsymbol{\delta}_i,$$

o que traduz que  $\hat{\boldsymbol{\delta}}_{i,OLS}$  não é um estimador centrado para  $\boldsymbol{\delta}_i$ .

Por outro lado, para estudar a consistência do estimador  $\hat{\boldsymbol{\delta}}_{i,OLS}$ , interessa escrever  $\hat{\boldsymbol{\delta}}_{i,OLS}$  noutra forma. Por (3.3.6),

$$\begin{aligned}
\hat{\boldsymbol{\delta}}_{i,OLS} &= \boldsymbol{\delta}_i + (\mathbf{Z}'_i \mathbf{Z}_i)^{-1} \mathbf{Z}'_i \mathbf{e}_i \\
&= \boldsymbol{\delta}_i + \left[ \begin{bmatrix} \mathbf{Y}'_i \\ \mathbf{X}'_i \end{bmatrix} \begin{bmatrix} \mathbf{Y}_i & \mathbf{X}_i \end{bmatrix} \right]^{-1} \begin{bmatrix} \mathbf{Y}'_i \\ \mathbf{X}'_i \end{bmatrix} \mathbf{e}_i \\
&= \boldsymbol{\delta}_i + \begin{bmatrix} \mathbf{Y}'_i \mathbf{Y}_i & \mathbf{Y}'_i \mathbf{X}_i \\ \mathbf{X}'_i \mathbf{Y}_i & \mathbf{X}'_i \mathbf{X}_i \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Y}'_i \mathbf{e}_i \\ \mathbf{X}'_i \mathbf{e}_i \end{bmatrix}.
\end{aligned}$$

Calculando o limite em probabilidade de  $\hat{\boldsymbol{\delta}}_{i,OLS}$  verifica-se que a segunda parcela tem limite não nulo (ver Apêndice A.2.1), o que traduz que o estimador  $\hat{\boldsymbol{\delta}}_{i,OLS}$  não é consistente para  $\boldsymbol{\delta}_i$ .

Note-se que isto acontece por se estar a utilizar o método dos mínimos quadrados, sem que estejam reunidas as hipóteses necessárias, uma vez que na equação  $\mathbf{y}_i = \mathbf{Z}_i \boldsymbol{\delta}_i + \mathbf{e}_i$  os regressores  $\mathbf{Z}_i = \begin{bmatrix} \mathbf{Y}_i & \mathbf{X}_i \end{bmatrix}$  e os erros aleatórios  $\mathbf{e}_i$  são correlacionados. De facto, uma vez que

$$p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{Y}'_i \mathbf{e}_i) = p \lim_{T \rightarrow \infty} \left[ \frac{1}{T} \sum_{t=1}^T y_{ti} e_{it} \right] \neq 0,$$

segue-se que

$$\begin{aligned}
 p \lim_{T \rightarrow \infty} \left[ \frac{\mathbf{Z}'_i \mathbf{e}_i}{T} \right] &= p \lim_{T \rightarrow \infty} \frac{1}{T} \left( \begin{bmatrix} \mathbf{Y}_i & \mathbf{X}_i \end{bmatrix}' \mathbf{e}_i \right) \\
 &= p \lim_{T \rightarrow \infty} \frac{1}{T} \left( \begin{bmatrix} \mathbf{Y}'_i \\ \mathbf{X}'_i \end{bmatrix} \mathbf{e}_i \right) \\
 &= p \lim_{T \rightarrow \infty} \frac{1}{T} \begin{bmatrix} \mathbf{Y}'_i \mathbf{e}_i \\ \mathbf{X}'_i \mathbf{e}_i \end{bmatrix} \neq \mathbf{0}.
 \end{aligned}$$

### 3.3.2 Estimador de variáveis instrumentais

Um método que permite ultrapassar a inconsistência dos estimadores dos parâmetros estruturais obtidos por *OLS* é o Método das Variáveis Instrumentais. Este método consiste em definir um conjunto de variáveis que sejam fortemente correlacionadas com os regressores, mas que não sejam correlacionadas com o erro aleatório da equação. Tais variáveis são chamadas variáveis instrumentais e podem ser construídas artificialmente ou podem surgir de forma natural em determinadas aplicações.

**Definição 3.3.1.** O estimador de variáveis instrumentais (EVI) para  $\delta_i$  é definido por

$$\hat{\delta}_{i,VI} = (\mathbf{W}'_i \mathbf{Z}_i)^{-1} \mathbf{W}'_i \mathbf{y}_i, \quad (3.3.7)$$

onde  $\mathbf{W}_i$  são as variáveis instrumentais para  $\mathbf{Z}_i$ , ou seja  $\mathbf{W}_i$  é uma matriz  $(T \times (M_i + k_i))$  que verifica as condições:

$$p \lim_{T \rightarrow \infty} \left[ \frac{\mathbf{W}'_i \mathbf{Z}_i}{T} \right] = \Sigma_{W_i Z_i}, \quad (3.3.8)$$

$$p \lim_{T \rightarrow \infty} \left[ \frac{\mathbf{W}'_i \mathbf{e}_i}{T} \right] = \mathbf{0}, \quad (3.3.9)$$

$$p \lim_{T \rightarrow \infty} \left[ \frac{\mathbf{W}'_i \mathbf{W}_i}{T} \right] = \Sigma_{W_i W_i}, \quad (3.3.10)$$

onde  $\Sigma_{W_i Z_i}$  é uma matriz finita e não singular e  $\Sigma_{W_i W_i}$  é uma matriz finita e definida positiva.  $\square$

Para obter este estimador, retome-se a equação  $i$  na forma estrutural definida em (2.3.4),  $\mathbf{y}_i = \mathbf{Z}_i \delta_i + \mathbf{e}_i$ , e multipliquem-se ambos os membros da equação  $i$  à esquerda



por  $\mathbf{W}'_i$ , com  $\mathbf{W}_i$  nas condições definidas de (3.3.8) a (3.3.10). Daí resulta a equação

$$\mathbf{W}'_i \mathbf{y}_i = \mathbf{W}'_i \mathbf{Z}_i \boldsymbol{\delta}_i + \mathbf{W}'_i \mathbf{e}_i. \quad (3.3.11)$$

Dividindo ambos os membros de (3.3.11) por  $T$  e calculando o limite em probabilidade, pode obter-se  $\boldsymbol{\delta}_i$ ,

$$\begin{aligned} p \lim_{T \rightarrow \infty} \frac{\mathbf{W}'_i \mathbf{y}_i}{T} &= p \lim_{T \rightarrow \infty} \frac{\mathbf{W}'_i \mathbf{Z}_i}{T} \boldsymbol{\delta}_i + p \lim_{T \rightarrow \infty} \frac{\mathbf{W}'_i \mathbf{e}_i}{T} \\ &\Leftrightarrow \Sigma_{W_i y_i} = \Sigma_{W_i Z_i} \boldsymbol{\delta}_i, \text{ por (3.3.8) e (3.3.9)} \\ &\Leftrightarrow \boldsymbol{\delta}_i = \Sigma_{W_i Z_i}^{-1} \Sigma_{W_i y_i}. \end{aligned} \quad (3.3.12)$$

Substituindo os momentos populacionais,  $\Sigma_{W_i y_i}$  e  $\Sigma_{W_i Z_i}$ , pelos correspondentes momentos amostrais,  $\mathbf{W}'_i \mathbf{y}_i$  e  $\mathbf{W}'_i \mathbf{Z}_i$ , em (3.3.12) obtém-se  $\hat{\boldsymbol{\delta}}_{i,VI} = (\mathbf{W}'_i \mathbf{Z}_i)^{-1} \mathbf{W}'_i \mathbf{y}_i$ , que é o estimador de variáveis instrumentais definido em (3.3.7).

O estimador de variáveis instrumentais goza das seguintes propriedades, cujas demonstrações se encontram no Apêndice A.2.2.

PVI.1 O estimador de variáveis instrumentais,  $\hat{\boldsymbol{\delta}}_{i,VI}$  é um estimador consistente para  $\boldsymbol{\delta}_i$ .

PVI.2 Se  $\frac{\mathbf{W}'_i \mathbf{e}_i}{\sqrt{T}} \dot{\sim} N(\mathbf{0}, \sigma_i^2 \Sigma_{W_i W_i})$ , então

$$\sqrt{T} \left( \hat{\boldsymbol{\delta}}_{i,VI} - \boldsymbol{\delta}_i \right) \dot{\sim} N \left( \mathbf{0}, \sigma_i^2 \left[ \Sigma_{W_i Z_i}^{-1} \Sigma_{W_i W_i} \Sigma_{Z_i W_i}^{-1} \right] \right).$$

PVI.3 A matriz de covariâncias assintótica de  $\hat{\boldsymbol{\delta}}_{i,VI}$  é

$$VA \left[ \hat{\boldsymbol{\delta}}_{i,VI} \right] = \frac{\sigma_i^2}{T} \Sigma_{W_i Z_i}^{-1} \Sigma_{W_i W_i} \Sigma_{Z_i W_i}^{-1},$$

e um estimador consistente para  $\sigma_i^2$  é  $\hat{\sigma}_i^2 = \frac{1}{T} \left[ \left( \mathbf{y}_i - \mathbf{Z}_i \hat{\boldsymbol{\delta}}_{i,VI} \right)' \left( \mathbf{y}_i - \mathbf{Z}_i \hat{\boldsymbol{\delta}}_{i,VI} \right) \right]$ .

Note-se que o número de variáveis instrumentais consideradas para a equação  $i$  deve ser igual ao número de regressores dessa equação, ou seja, o número de colunas de  $\mathbf{W}_i$  deve ser igual ao número de colunas de  $\mathbf{Z}_i$ , para assegurar que a matriz  $\mathbf{W}'_i \mathbf{Z}_i$  seja uma matriz quadrada e que exista a respectiva inversa.

### 3.3.3 Estimador de mínimos quadrados indirectos

Como se viu, embora não seja possível obter estimadores consistentes para os parâmetros estruturais aplicando directamente o método de mínimos quadrados a cada equação

da forma (2.3.4), os parâmetros da forma reduzida podem ser estimados de forma consistente pelo método dos mínimos quadrados. Admitindo que o actual objectivo é estimar os parâmetros estruturais do modelo e que estes se relacionam com os parâmetros da forma reduzida através de um sistema de equações, podem estimar-se primeiro os parâmetros da forma reduzida, por *OLS*, e, a partir destes, obter os estimadores dos parâmetros estruturais, resolvendo o referido sistema. O estimador assim obtido chama-se estimador de mínimos quadrados indirectos (*ILS*), e será representado por  $\hat{\delta}_{i,ILS}$ .

Embora seja um processo bastante intuitivo, nem sempre permite obter estimadores únicos dos parâmetros estruturais. Esta questão já foi referida e está relacionada com o problema da identificabilidade das equações.

De facto, o estudo da identificabilidade resume-se a averiguar se existe informação suficiente no modelo, ou seja, se existem, *à priori*, restrições suficientes, que permitam utilizar o estimador *ILS* para estimar os parâmetros estruturais de uma equação.

Assim, para uma equação particular  $i$  de um *SEM*, podem surgir as seguintes situações:

- Se a equação  $i$  for não identificada, não é possível obter os parâmetros estruturais dessa equação a partir dos parâmetros da forma reduzida (caso em que o sistema é algebricamente impossível). Neste caso não é possível obter estimadores consistentes dos parâmetros estruturais.
- Se a equação  $i$  for exactamente identificada, os parâmetros estruturais podem expressar-se de forma única em função dos parâmetros da forma reduzida (caso em que o sistema tem tantas equações como parâmetros desconhecidos e tem solução única). Neste caso é possível obter estimadores consistentes dos parâmetros estruturais.
- Se a equação  $i$  for sobreidentificada, há várias formas de exprimir os parâmetros estruturais em função dos parâmetros da forma reduzida (caso em que o sistema tem mais equações do que parâmetros desconhecidos e admite uma infinidade de soluções). Neste caso é possível obter estimadores consistentes dos parâmetros estruturais, mas esses estimadores não são únicos.

Pelo que se acabou de referir, só é possível proceder à estimação de parâmetros estruturais de equações identificadas. Assim, tendo em conta que o principal objectivo do presente capítulo é a estimação dos parâmetros estruturais de um *SEM*, admite-se no seguimento que todas as equações do *SEM* são identificadas.

Nessas condições, para obter o estimador *ILS* dos parâmetros de uma equação identificada, parte-se da forma reduzida dessa equação e estimam-se os respectivos parâmetros por *OLS*. Como se viu na **Subsecção 3.3.1**, este método de estimação é adequado, uma vez que na forma reduzida todos os regressores são variáveis exógenas e estas, por hipótese, são não correlacionadas com os erros da forma estrutural. Seguidamente, resolve-se o sistema de equações que relacionam os parâmetros da forma estrutural com os da forma reduzida, nomeadamente, as equações

$$\mathbf{B} = -\mathbf{\Pi}\mathbf{\Gamma} \text{ e } \mathbf{\tilde{\Sigma}} = \mathbf{\Gamma}'\mathbf{\tilde{\Sigma}}^v\mathbf{\Gamma}, \quad (3.3.13)$$

obtidas a partir de (2.2.12) e (2.2.15), respectivamente. Obtêm-se, assim, os estimadores dos parâmetros estruturais da equação.

Se a equação for sobreidentificada, obtêm-se vários estimadores *ILS* distintos para alguns parâmetros estruturais e não há nenhum critério objectivo que permita escolher entre eles. Se a equação for exactamente identificada, são obtidos estimadores únicos dos parâmetros estruturais, pelo processo que se descreve de seguida.

Considere-se a equação  $i$  do modelo na forma reduzida,  $\mathbf{y}_i = \mathbf{X}\boldsymbol{\pi}_i + \mathbf{v}_i$ . Como se viu na **Subsecção 3.3.1**, o estimador  $\hat{\boldsymbol{\pi}}_i = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_i$  é um estimador consistente e centrado para  $\boldsymbol{\pi}_i$ . Repetindo o procedimento para todas as equações, é possível obter

$$\hat{\mathbf{\Pi}} = \begin{bmatrix} \hat{\boldsymbol{\pi}}_1 & \cdots & \hat{\boldsymbol{\pi}}_M \end{bmatrix}.$$

A partir das estimativas  $\hat{\boldsymbol{\pi}}_i$ , podem obter-se os resíduos da forma reduzida, definidos por

$$\hat{\mathbf{v}}_i = \mathbf{y}_i - \mathbf{X}\hat{\boldsymbol{\pi}}_i$$

e, com base nestes, pode estimar-se a matriz  $\mathbf{\tilde{\Sigma}}^v$  de forma consistente, por

$$\widehat{\mathbf{\tilde{\Sigma}}^v} = [\hat{w}_{ij}], \text{ onde } \hat{w}_{ij} = \frac{\hat{\mathbf{v}}_i'\hat{\mathbf{v}}_j}{T-k}, i, j = 1, \dots, M.$$

Finalmente, substituindo  $\hat{\mathbf{\Pi}}$  e  $\widehat{\mathbf{\tilde{\Sigma}}^v}$ , nas equações (3.3.13), obtêm-se estimativas dos parâmetros estruturais, para as equações identificadas.

A forma do estimador *ILS* pode ser expressa em termos de  $\Delta_i = \begin{bmatrix} \Gamma_i' & \mathbf{B}_i' \end{bmatrix}'$ .

Note-se que  $\Delta_i \neq \delta_i$ , uma vez que

$$\Delta_i = \begin{bmatrix} \Gamma_i \\ \mathbf{B}_i \end{bmatrix}_{((M+k) \times 1)} = \begin{bmatrix} -1 \\ \gamma_i \\ \mathbf{0} \\ \beta_i \\ \mathbf{0} \end{bmatrix} \quad (3.3.14)$$

e  $\delta_i = \begin{bmatrix} \gamma_i \\ \beta_i \end{bmatrix}_{((M^*+k^*) \times 1)}$ .

Pelo referido no estudo da identificabilidade da equação  $i$ , nomeadamente em (2.3.8), tem-se

$$\begin{bmatrix} \left( \begin{array}{cc} \Pi & \mathbf{I}_k \end{array} \right) \\ \mathbf{R}_i \end{bmatrix} \Delta_i = \mathbf{0},$$

e o estimador *ILS* para  $\Delta_i$  resultará como solução da equação

$$\begin{bmatrix} \left( \begin{array}{cc} \hat{\Pi} & \mathbf{I}_k \end{array} \right) \\ \mathbf{R}_i \end{bmatrix} \Delta_i = \mathbf{0},$$

onde  $\hat{\Pi}$  é o estimador obtido por *OLS* para os parâmetros da forma reduzida. Designando por  $\hat{\mathbf{A}}$  a matriz

$$\hat{\mathbf{A}} = \begin{bmatrix} \left( \begin{array}{cc} \hat{\Pi} & \mathbf{I}_k \end{array} \right) \\ \mathbf{R}_i \end{bmatrix}_{(M+k-1) \times (M+k)}, \quad (3.3.15)$$

então o estimador *ILS* para  $\Delta_i$  é solução da equação matricial

$$\hat{\mathbf{A}} \Delta_i = \mathbf{0}, \quad (3.3.16)$$

e é da forma  $\Delta_i = \begin{bmatrix} -1 \\ \Delta_i^* \end{bmatrix}$ .

Resumindo, para estimar os parâmetros estruturais de uma equação identificada por *ILS*, devem seguir-se os seguintes passos:

1. Passar o *SEM* à forma reduzida.
2. Estimar os parâmetros da forma reduzida, por *OLS*.

3. A partir destes, e recorrendo a (3.3.13), obter os estimadores dos parâmetros estruturais.

O estimador *ILS* será denotado por  $\hat{\Delta}_i$  e goza das propriedades que se apresentam de seguida, cujas demonstrações podem ser vistas em Schmidt (1976).

PILS.1 O estimador *ILS* é consistente, isto é,  $p \lim_{T \rightarrow \infty} \hat{\Delta}_i = \Delta_i$ .

PILS.2 Se a equação *i* for exactamente identificada, o estimador *ILS* é único.

PILS.3 Se a equação *i* for sobreidentificada, o estimador *ILS* não é único.

PILS.4 O estimador *ILS* é um estimador de variáveis instrumentais.

Repare-se que o estimador *ILS* é superior ao estimador de *OLS* para estimar os parâmetros estruturais, pois é consistente. No entanto, o estimador *ILS* não é eficiente para estimar os parâmetros de equações sobreidentificadas, uma vez que, neste caso, despreza parte da informação amostral.

Tendo em consideração que o estimador *ILS* é obtido como solução da equação (3.3.16) e efectuando alguns cálculos matriciais, é possível obter para o estimador *ILS* de  $\delta_i$ , a forma seguinte,

$$\hat{\delta}_{i,ILS} = \left[ (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z}_i \right]^{-1} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}_i. \quad (3.3.17)$$

Apresenta-se de seguida um estimador com melhores propriedades do que os dois anteriores, na medida em que é consistente para estimar os parâmetros estruturais de qualquer equação identificada e é assintoticamente eficiente.

### 3.3.4 Estimador de mínimos quadrados generalizados

Para obter o estimador de mínimos quadrados generalizados (*GLS* de *Generalized Least Squares*), considere-se de novo a equação *i* na forma estrutural,  $\mathbf{y}_i = \mathbf{Z}_i\delta_i + \mathbf{e}_i$ , e proceda-se à transformação

$$\mathbf{X}'\mathbf{y}_i = \mathbf{X}'\mathbf{Z}_i\delta_i + \mathbf{X}'\mathbf{e}_i.$$

Esta transformação permitirá obter um estimador consistente para  $\delta_i$ . Para isso, recorde-se que por HFE.4 se tem que a matriz  $\Sigma_{\mathbf{XX}} = p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{X}'\mathbf{X})$  existe e é finita. Mas, usando a decomposição definida na Tabela 2.1, pode obter-se a forma

$$\begin{aligned} \Sigma_{\mathbf{XX}} &= p \lim_{T \rightarrow \infty} \frac{1}{T} \begin{bmatrix} \mathbf{X}'_i \mathbf{X}_i & \mathbf{X}'_i \mathbf{X}_i^* \\ \mathbf{X}_i^{*'} \mathbf{X}_i & \mathbf{X}_i^{*'} \mathbf{X}_i^* \end{bmatrix} \\ \Leftrightarrow \Sigma_{\mathbf{XX}} &= \begin{bmatrix} (\Sigma_{\mathbf{XX}})_{11} & (\Sigma_{\mathbf{XX}})_{12} \\ (\Sigma_{\mathbf{XX}})_{21} & (\Sigma_{\mathbf{XX}})_{22} \end{bmatrix} \\ \Leftrightarrow \Sigma_{\mathbf{XX}} &= \begin{bmatrix} (\Sigma_{\mathbf{XX}})_{.1} & (\Sigma_{\mathbf{XX}})_{.2} \end{bmatrix}. \end{aligned} \quad (3.3.18)$$

Pode daqui concluir-se que

$$p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{X}'\mathbf{X}_i) = (\Sigma_{\mathbf{XX}})_{.1}. \quad (3.3.19)$$

Por outro lado, pela **Proposição 2.2.3**, tem-se que  $p \lim_{T \rightarrow \infty} (\mathbf{X}'\mathbf{Y}) = \Sigma_{\mathbf{XX}} \Pi$ . Desenvolvendo esta forma matricial, de acordo com a decomposição de matrizes definida na Tabela 2.1, vem

$$\begin{aligned} \Pi' \Sigma_{\mathbf{XX}} &= \begin{bmatrix} \pi'_{11} & \pi'_{21} \\ \pi'_{12} & \pi'_{22} \\ \pi'_{13} & \pi'_{23} \end{bmatrix} \begin{bmatrix} (\Sigma_{\mathbf{XX}})_{11} & (\Sigma_{\mathbf{XX}})_{12} \\ (\Sigma_{\mathbf{XX}})_{21} & (\Sigma_{\mathbf{XX}})_{22} \end{bmatrix} \\ &= \begin{bmatrix} \pi'_{.1} (\Sigma_{\mathbf{XX}})_{.1} & \pi'_{.1} (\Sigma_{\mathbf{XX}})_{.2} \\ \pi'_{.2} (\Sigma_{\mathbf{XX}})_{.1} & \pi'_{.2} (\Sigma_{\mathbf{XX}})_{.2} \\ \pi'_{.3} (\Sigma_{\mathbf{XX}})_{.1} & \pi'_{.3} (\Sigma_{\mathbf{XX}})_{.2} \end{bmatrix}, \end{aligned}$$

e

$$p \lim_{T \rightarrow \infty} (\mathbf{X}'\mathbf{Y}) = p \lim_{T \rightarrow \infty} \frac{1}{T} \begin{bmatrix} \mathbf{y}'_i \mathbf{X}_i & \mathbf{y}'_i \mathbf{X}_i^* \\ \mathbf{Y}'_i \mathbf{X}_i & \mathbf{Y}'_i \mathbf{X}_i^* \\ \mathbf{Y}_i^{*'} \mathbf{X}_i & \mathbf{Y}_i^{*'} \mathbf{X}_i^* \end{bmatrix}.$$

Deste modo,

$$\begin{aligned} p \lim_{T \rightarrow \infty} (\mathbf{X}'\mathbf{Y}) &= \Sigma_{\mathbf{XX}} \Pi \\ \Leftrightarrow p \lim_{T \rightarrow \infty} \frac{1}{T} \begin{bmatrix} \mathbf{y}'_i \mathbf{X}_i & \mathbf{y}'_i \mathbf{X}_i^* \\ \mathbf{Y}'_i \mathbf{X}_i & \mathbf{Y}'_i \mathbf{X}_i^* \\ \mathbf{Y}_i^{*'} \mathbf{X}_i & \mathbf{Y}_i^{*'} \mathbf{X}_i^* \end{bmatrix} &= \begin{bmatrix} \pi'_{.1} (\Sigma_{\mathbf{XX}})_{.1} & \pi'_{.1} (\Sigma_{\mathbf{XX}})_{.2} \\ \pi'_{.2} (\Sigma_{\mathbf{XX}})_{.1} & \pi'_{.2} (\Sigma_{\mathbf{XX}})_{.2} \\ \pi'_{.3} (\Sigma_{\mathbf{XX}})_{.1} & \pi'_{.3} (\Sigma_{\mathbf{XX}})_{.2} \end{bmatrix}, \end{aligned}$$

donde se conclui que

$$p \lim_{T \rightarrow \infty} (\mathbf{Y}'_i \mathbf{X}) = \pi'_{.2} \Sigma_{\mathbf{XX}} \Leftrightarrow p \lim_{T \rightarrow \infty} (\mathbf{X}'\mathbf{Y}_i) = \Sigma_{\mathbf{XX}} \pi_{.2}. \quad (3.3.20)$$

Assim, como  $\mathbf{Z}_i = \begin{bmatrix} \mathbf{Y}_i & \mathbf{X}_i \end{bmatrix}$ , vem que o limite

$$\begin{aligned} p \lim_{T \rightarrow \infty} \left[ \frac{\mathbf{X}'\mathbf{Z}_i}{T} \right] &= p \lim_{T \rightarrow \infty} \begin{bmatrix} \frac{\mathbf{X}'\mathbf{Y}_i}{T} & \frac{\mathbf{X}'\mathbf{X}_i}{T} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Y}} & (\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}})_{.1} \end{bmatrix}, \end{aligned}$$

o que traduz que  $p \lim_{T \rightarrow \infty} \left[ \frac{\mathbf{X}'\mathbf{Z}_i}{T} \right]$  existe e é finito, o qual será designado por  $\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Z}_i}$ . Assim, assintoticamente,  $\mathbf{X}'\mathbf{Z}_i$  é uma matriz constante, pelo que pode considerar-se uma matriz de regressores não correlacionada com os erros aleatórios  $\mathbf{X}'\mathbf{e}_i$ . Para obter o estimador de mínimos quadrados generalizados é necessário estimar a variância dos erros  $\mathbf{X}'\mathbf{e}_i$ ,

$$V[\mathbf{X}'\mathbf{e}_i] = E[\mathbf{X}'\mathbf{e}_i\mathbf{e}_i'\mathbf{X}] = \sigma_i^2 E[\mathbf{X}'\mathbf{X}].$$

A matriz  $E[\mathbf{X}'\mathbf{X}]$  é desconhecida, mas pode ser estimada por  $(\mathbf{X}'\mathbf{X})$  e o valor de  $\sigma_i^2$ , que é desconhecido e não nulo, é simplificado nos cálculos intermédios, permitindo obter

$$\begin{aligned} \hat{\boldsymbol{\delta}}_{i, GLS} &= \left[ (\mathbf{X}'\mathbf{Z}_i)' (\sigma_i^2 \mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z}_i \right]^{-1} (\mathbf{X}'\mathbf{Z}_i)' (\sigma_i^2 \mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}_i \\ &= \left[ \mathbf{Z}_i'\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z}_i \right]^{-1} \mathbf{Z}_i'\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}_i, \end{aligned} \quad (3.3.21)$$

admitindo que existe a inversa da matriz  $[\mathbf{Z}_i'\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z}_i]$ . Trata-se de uma matriz quadrada de dimensão  $(M_i + k_i - 1)$ ; logo, para existir inversa tem que se verificar a condição  $(M_i + k_i - 1) \leq k$ . Tendo em conta que

$$\text{rank}(\mathbf{Z}_i'\mathbf{X}) \leq \min \left\{ \text{rank}(\mathbf{Z}_i'), \underbrace{\text{rank}(\mathbf{X})}_k \right\},$$

vem que

$$\begin{aligned} (M_i + k_i - 1) \leq k &\Leftrightarrow (M_i - 1) \leq k - k_i \\ &\Leftrightarrow (M_i - 1) \leq k_i^*, \end{aligned}$$

onde  $k_i^*$  é o número de variáveis exógenas com coeficiente nulo na equação  $i$ , ou, dito de outro modo, é o número de restrições de exclusão sobre variáveis exógenas da equação  $i$ . Note-se que esta é a condição necessária para a identificabilidade da equação  $i$ , referida em (2.3.11), pelo que, o estimador  $\hat{\boldsymbol{\delta}}_{i, GLS}$  fica sempre definido para uma equação  $i$  que seja identificada.

O estimador de mínimos quadrados generalizados pode ser expresso em várias formas, admitindo assim diferentes interpretações. Em particular, o estimador *GLS* pode ser relacionado com o estimador de mínimos quadrados indirectos, estudado na subsecção anterior. No caso da equação  $i$  ser exactamente identificada, ambos os estimadores coincidem, como se passa a verificar. De facto, nesse caso,  $\mathbf{X}'\mathbf{Z}_i$  é uma matriz quadrada de ordem  $k$  e pode efectuar-se a seguinte simplificação,

$$\begin{aligned}\hat{\delta}_{i, GLS} &= \left[ \mathbf{Z}_i' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z}_i \right]^{-1} \mathbf{Z}_i' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}_i \\ &= (\mathbf{X}' \mathbf{Z}_i)^{-1} (\mathbf{X}' \mathbf{X}) (\mathbf{Z}_i' \mathbf{X})^{-1} \mathbf{Z}_i' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}_i \\ &= (\mathbf{X}' \mathbf{Z}_i)^{-1} \mathbf{X}' \mathbf{y}_i.\end{aligned}$$

Por outro lado, a partir de (3.3.17) tem-se  $\hat{\delta}_{i, ILS} = [(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z}_i]^{-1} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}_i$  que, pelos mesmos motivos, se simplifica na mesma expressão, confirmando que  $\hat{\delta}_{i, GLS} = \hat{\delta}_{i, ILS}$ .

O estimador *GLS* goza das seguintes propriedades, cujas demonstrações podem ser consultadas em Judge *et al.* (1988):

PGLS.1 O estimador  $\hat{\delta}_{i, GLS}$  é um estimador de variáveis instrumentais.

PGLS.2 O estimador  $\hat{\delta}_{i, GLS}$  é um estimador consistente para  $\delta_i$  e tem-se

$$\sqrt{T} \left( \hat{\delta}_{i, GLS} - \delta_i \right) \overset{\circ}{\sim} N \left( \mathbf{0}, \sigma_i^2 \left[ \Sigma_{\mathbf{XZ}_i}^{-1} \Sigma_{\mathbf{XX}} \Sigma_{\mathbf{Z}_i \mathbf{X}}^{-1} \right] \right).$$

Um estimador consistente para esta matriz de covariâncias assintótica é o seguinte

$$\widehat{VA} \left[ \hat{\delta}_{i, GLS} \right] = \hat{\sigma}_i^2 \left[ (\mathbf{Z}_i' \mathbf{X}) (\mathbf{X}' \mathbf{X})^{-1} (\mathbf{X}' \mathbf{Z}_i) \right]^{-1},$$

onde  $\hat{\sigma}_i^2$  é um estimador consistente para  $\sigma_i^2$ , sendo habitualmente usado

$$\hat{\sigma}_{i, GLS}^2 = \frac{1}{T} \left[ \left( \mathbf{y}_i - \mathbf{Z}_i \hat{\delta}_{i, GLS} \right)' \left( \mathbf{y}_i - \mathbf{Z}_i \hat{\delta}_{i, GLS} \right) \right].$$

### 3.3.5 Estimador de mínimos quadrados em duas fases

Ao estimar os parâmetros estruturais de uma equação sobreidentificada, o estimador *ILS* fornece várias estimativas, conforme se viu em PILS.3. O estimador de mínimos quadrados em duas fases (*2SLS*) surgiu para estimar este tipo de equações, embora também possa ser utilizado para estimar equações exactamente identificadas e, ao



contrário do estimador *ILS*, permite obter uma estimativa única para o vector de parâmetros de uma equação sobreidentificada. A ideia base deste estimador consiste em substituir as variáveis endógenas, que são regressores em algumas equações, por combinações lineares das variáveis exógenas e usar estas combinações lineares como regressores, no lugar das variáveis endógenas. Na prática, a obtenção deste estimador passa, numa primeira fase, por estimar os parâmetros da forma reduzida por *OLS* e obter os valores predictos  $\hat{\mathbf{Y}}_i$  e, numa segunda fase, por estimar  $\boldsymbol{\delta}_i$  por *OLS* na equação  $i$  que se obtém de (2.3.3), depois de substituir  $\mathbf{Y}_i$  por  $\hat{\mathbf{Y}}_i$ .

Para deduzir a expressão do estimador de mínimos quadrados em duas fases (*2SLS*), considera-se a forma reduzida do modelo, definida em (2.2.11) e procede-se à partição das matrizes da forma reduzida, de forma concordante com a definida na Tabela 2.1. Assim, vem

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\boldsymbol{\Pi} + \mathbf{V} \\ \Leftrightarrow \underbrace{\begin{bmatrix} \mathbf{y}_i & \mathbf{Y}_i & \mathbf{Y}_i^* \end{bmatrix}}_{(T \times (1+(M_i-1)+M_i^*))} &= \underbrace{\begin{bmatrix} \mathbf{X}_i & \mathbf{X}_i^* \end{bmatrix}}_{(T \times (k_i+k_i^*))} \underbrace{\begin{bmatrix} \boldsymbol{\pi}_{11} & \boldsymbol{\pi}_{12} & \boldsymbol{\pi}_{13} \\ \boldsymbol{\pi}_{21} & \boldsymbol{\pi}_{22} & \boldsymbol{\pi}_{23} \end{bmatrix}}_{((k_i+k_i^*) \times (1+(M_i-1)+M_i^*))} + \underbrace{\begin{bmatrix} \mathbf{v}_i & \mathbf{V}_i & \mathbf{V}_i^* \end{bmatrix}}_{(T \times (1+(M_i-1)+M_i^*))} \\ & \quad (3.3.22) \\ \Leftrightarrow \begin{bmatrix} \mathbf{y}_i & \mathbf{Y}_i & \mathbf{Y}_i^* \end{bmatrix} &= \begin{bmatrix} \mathbf{X}_i & \mathbf{X}_i^* \end{bmatrix} \begin{bmatrix} \boldsymbol{\pi}_{.1} & \boldsymbol{\pi}_{.2} & \boldsymbol{\pi}_{.3} \end{bmatrix} + \begin{bmatrix} \mathbf{v}_i & \mathbf{V}_i & \mathbf{V}_i^* \end{bmatrix}, \end{aligned}$$

onde  $\boldsymbol{\pi}_{.i}$  traduz a coluna  $i$  da matriz  $\boldsymbol{\Pi}$ . Pode retirar-se desta última forma que

$$\mathbf{Y}_i = \mathbf{X}\boldsymbol{\pi}_{.2} + \mathbf{V}_i. \quad (3.3.23)$$

De acordo com este método, na primeira fase, estimam-se os parâmetros da forma reduzida por *OLS*; ou seja, calcula-se  $\hat{\boldsymbol{\Pi}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$ , que se pode reescrever usando a partição definida em (3.3.22), do seguinte modo

$$\begin{aligned} \begin{bmatrix} \hat{\boldsymbol{\pi}}_{.1} & \hat{\boldsymbol{\pi}}_{.2} & \hat{\boldsymbol{\pi}}_{.3} \end{bmatrix} &= (\mathbf{X}'\mathbf{X})^{-1} \begin{bmatrix} \mathbf{X}_i & \mathbf{X}_i^* \end{bmatrix}' \begin{bmatrix} \mathbf{y}_i & \mathbf{Y}_i & \mathbf{Y}_i^* \end{bmatrix} \\ \Leftrightarrow \begin{bmatrix} \hat{\boldsymbol{\pi}}_{.1} & \hat{\boldsymbol{\pi}}_{.2} & \hat{\boldsymbol{\pi}}_{.3} \end{bmatrix} &= (\mathbf{X}'\mathbf{X})^{-1} \begin{bmatrix} \mathbf{X}_i' \mathbf{y}_i & \mathbf{X}_i' \mathbf{Y}_i & \mathbf{X}_i' \mathbf{Y}_i^* \\ (\mathbf{X}_i^*)' \mathbf{y}_i & (\mathbf{X}_i^*)' \mathbf{Y}_i & (\mathbf{X}_i^*)' \mathbf{Y}_i^* \end{bmatrix}. \end{aligned}$$

Donde resulta que  $\hat{\boldsymbol{\pi}}_{.2} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}_i$ . Ao substituir  $\boldsymbol{\pi}_{.2}$  por  $\hat{\boldsymbol{\pi}}_{.2}$  em (3.3.23) obtém-se

uma forma para os valores preditos,  $\hat{\mathbf{Y}}_i$

$$\begin{aligned}
\hat{\mathbf{Y}}_i &= \mathbf{X}\hat{\boldsymbol{\pi}}_{.2} \Leftrightarrow \hat{\mathbf{Y}}_i = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \begin{bmatrix} \mathbf{X}_i'\mathbf{Y}_i \\ (\mathbf{X}_i^*)'\mathbf{Y}_i \end{bmatrix} \\
&\Leftrightarrow \hat{\mathbf{Y}}_i = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \begin{bmatrix} \mathbf{X}_i' \\ (\mathbf{X}_i^*)' \end{bmatrix} \mathbf{Y}_i \\
&\Leftrightarrow \hat{\mathbf{Y}}_i = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \begin{bmatrix} \mathbf{X}_i & \mathbf{X}_i^* \end{bmatrix}' \mathbf{Y}_i \\
&\Leftrightarrow \hat{\mathbf{Y}}_i = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}_i.
\end{aligned} \tag{3.3.24}$$

De seguida, na segunda fase, estima-se  $\boldsymbol{\delta}_i$  por *OLS* a partir da equação que se obtém da forma (2.3.3), substituindo  $\mathbf{Y}_i$  por  $\hat{\mathbf{Y}}_i$  e fazendo  $\bar{\mathbf{e}}_i = \mathbf{e}_i + (\mathbf{Y}_i - \hat{\mathbf{Y}}_i)\boldsymbol{\gamma}_i$ , ou seja,

$$\begin{aligned}
\mathbf{y}_i &= \hat{\mathbf{Y}}_i\boldsymbol{\gamma}_i + \mathbf{X}_i\boldsymbol{\beta}_i + \bar{\mathbf{e}}_i \\
&\Leftrightarrow \mathbf{y}_i = \begin{bmatrix} \hat{\mathbf{Y}}_i & \mathbf{X}_i \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma}_i \\ \boldsymbol{\beta}_i \end{bmatrix} + \bar{\mathbf{e}}_i \\
&\Leftrightarrow \mathbf{y}_i = \begin{bmatrix} \hat{\mathbf{Y}}_i & \mathbf{X}_i \end{bmatrix} \boldsymbol{\delta}_i + \bar{\mathbf{e}}_i \\
&\Leftrightarrow \mathbf{y}_i = \hat{\mathbf{Z}}_i\boldsymbol{\delta}_i + \bar{\mathbf{e}}_i.
\end{aligned} \tag{3.3.25}$$

Deste modo, estimando  $\boldsymbol{\delta}_i$  por *OLS* no modelo (3.3.25), obtém-se a seguinte forma para o estimador *2SLS*,

$$\hat{\boldsymbol{\delta}}_{i,2SLS} = \left( \hat{\mathbf{Z}}_i'\hat{\mathbf{Z}}_i \right)^{-1} \hat{\mathbf{Z}}_i'\mathbf{y}_i, \text{ com } \hat{\mathbf{Z}}_i = \begin{bmatrix} \hat{\mathbf{Y}}_i & \mathbf{X}_i \end{bmatrix}, \tag{3.3.26}$$

ou seja,

$$\hat{\boldsymbol{\delta}}_{i,2SLS} = \left\{ \underbrace{\begin{bmatrix} \hat{\mathbf{Y}}_i & \mathbf{X}_i \end{bmatrix}' \begin{bmatrix} \hat{\mathbf{Y}}_i & \mathbf{X}_i \end{bmatrix}}_{((M_i+k_i-1) \times T) \quad (T \times (M_i+k_i-1))} \right\}^{-1} \begin{bmatrix} \hat{\mathbf{Y}}_i & \mathbf{X}_i \end{bmatrix}' \mathbf{y}_i. \tag{3.3.27}$$

Repare-se que a equação  $i$  na forma (3.3.25) verifica as hipóteses necessárias para aplicação do *OLS*, uma vez que entre os regressores,  $\begin{bmatrix} \hat{\mathbf{Y}}_i & \mathbf{X}_i \end{bmatrix}$ , já não se encontram variáveis endógenas, pelo que os regressores e os erros aleatórios do modelo são não correlacionados.

Repetindo este procedimento para cada uma das  $M$  equações do modelo na forma estrutural, podem obter-se estimadores *2SLS* para todos os parâmetros estruturais do modelo.

Para estudar as propriedades do estimador  $2SLS$  para  $\delta_i$  é importante notar que o estimador  $\hat{\delta}_{i,2SLS}$  pode ser representado noutras formas alternativas, de acordo com os desenvolvimentos seguintes. Partindo de (3.3.27), vem

$$\begin{aligned}\hat{\delta}_{i,2SLS} &= \left\{ \begin{bmatrix} \hat{\mathbf{Y}}'_i \\ \mathbf{X}'_i \end{bmatrix} \begin{bmatrix} \hat{\mathbf{Y}}_i & \mathbf{X}_i \end{bmatrix} \right\}^{-1} \begin{bmatrix} \hat{\mathbf{Y}}'_i \\ \mathbf{X}'_i \end{bmatrix} \mathbf{y}_i \\ \Leftrightarrow \hat{\delta}_{i,2SLS} &= \begin{bmatrix} \hat{\mathbf{Y}}'_i \hat{\mathbf{Y}}_i & \hat{\mathbf{Y}}'_i \mathbf{X}_i \\ \mathbf{X}'_i \hat{\mathbf{Y}}_i & \mathbf{X}'_i \mathbf{X}_i \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{Y}}'_i \mathbf{y}_i \\ \mathbf{X}'_i \mathbf{y}_i \end{bmatrix}\end{aligned}\quad (3.3.28)$$

$$\Leftrightarrow \hat{\delta}_{i,2SLS} = \left( \hat{\mathbf{Z}}'_i \mathbf{Z}_i \right)^{-1} \hat{\mathbf{Z}}'_i \mathbf{y}_i. \quad (3.3.29)$$

Esta forma do estimador também pode ser obtida partindo da forma (3.3.21) do estimador  $GLS$  para  $\delta_i$  e efectuando algumas simplificações, do seguinte modo,

$$\begin{aligned}\hat{\delta}_{i,GLS} &= \left[ \mathbf{Z}'_i \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z}_i \right]^{-1} \mathbf{Z}'_i \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}_i \\ &= \left\{ \begin{bmatrix} \mathbf{Y}_i & \mathbf{X}_i \end{bmatrix}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \begin{bmatrix} \mathbf{Y}_i & \mathbf{X}_i \end{bmatrix} \right\}^{-1} \begin{bmatrix} \mathbf{Y}_i & \mathbf{X}_i \end{bmatrix}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}_i \\ &= \underbrace{\begin{bmatrix} \mathbf{Y}'_i \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}_i & \mathbf{Y}'_i \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{X}_i \\ \mathbf{X}'_i \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}_i & \mathbf{X}'_i \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{X}_i \end{bmatrix}}_{\mathbf{A}_1}^{-1} \underbrace{\begin{bmatrix} \mathbf{Y}'_i \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}_i \\ \mathbf{X}'_i \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}_i \end{bmatrix}}_{\mathbf{A}_2}.\end{aligned}$$

As matrizes  $\mathbf{A}_1$  e  $\mathbf{A}_2$  simplificam-se com artifícios simples. Atendendo à expressão de  $\hat{\mathbf{Y}}_i$  referida em (3.3.24), vem que

$$\mathbf{A}_1 = \begin{bmatrix} \hat{\mathbf{Y}}'_i \hat{\mathbf{Y}}_i & \hat{\mathbf{Y}}'_i \mathbf{X}_i \\ \mathbf{X}'_i \hat{\mathbf{Y}}_i & \mathbf{X}'_i \mathbf{X}_i \end{bmatrix}^{-1} \text{ e } \mathbf{A}_2 = \begin{bmatrix} \hat{\mathbf{Y}}'_i \mathbf{y}_i \\ \mathbf{X}'_i \mathbf{y}_i \end{bmatrix},$$

logo

$$\hat{\delta}_{i,GLS} = \left( \hat{\mathbf{Z}}'_i \mathbf{Z}_i \right)^{-1} \hat{\mathbf{Z}}'_i \mathbf{y}_i = \hat{\delta}_{i,2SLS}. \quad (3.3.30)$$

Verifica-se assim que o estimador  $\hat{\delta}_{i,2SLS}$  pode ser considerado uma representação do estimador  $GLS$  para  $\delta_i$ .

Deste modo, as propriedades enunciadas para o estimador  $\hat{\delta}_{i,GLS}$  são válidas também para o estimador  $\hat{\delta}_{i,2SLS}$ . Para além dessas, podem ainda referir-se outras propriedades do estimador  $2SLS$ , que se apresentam de seguida e cujas demonstrações podem ser vistas em Schmidt (1976).

P2SLS.1 O estimador  $\hat{\delta}_{i,2SLS}$  pode ser apresentado na forma  $\hat{\delta}_{i,2SLS} = \left(\hat{\mathbf{Z}}_i' \mathbf{Z}_i\right)^{-1} \hat{\mathbf{Z}}_i' \mathbf{y}_i$ , referida em (3.4.13), evidenciando que é um estimador de variáveis instrumentais para  $\delta_i$ .

P2SLS.2 O estimador  $\hat{\delta}_{i,2SLS}$  é um estimador consistente para  $\delta_i$  e tem-se que

$$\sqrt{T} \left( \hat{\delta}_{i,2SLS} - \delta_i \right) \stackrel{\cdot}{\sim} N \left( \mathbf{0}, \sigma_i^2 \left[ \Sigma_{\mathbf{X}\mathbf{Z}_i}^{-1} \Sigma_{\mathbf{X}\mathbf{X}} \Sigma_{\mathbf{Z}_i\mathbf{X}}^{-1} \right] \right).$$

A matriz de covariâncias assintótica pode ser estimada de forma consistente por

$$\widehat{VA} \left[ \hat{\delta}_{i,2SLS} \right] = \hat{\sigma}_i^2 \left[ (\mathbf{Z}_i' \mathbf{X}) (\mathbf{X}' \mathbf{X})^{-1} (\mathbf{X}' \mathbf{Z}_i) \right]^{-1},$$

$$\text{onde } \hat{\sigma}_i^2 = \frac{1}{T} \left[ \left( \mathbf{y}_i - \mathbf{Z}_i \hat{\delta}_{i,2SLS} \right)' \left( \mathbf{y}_i - \mathbf{Z}_i \hat{\delta}_{i,2SLS} \right) \right].$$

P2SLS.3 O estimador  $\hat{\delta}_{i,2SLS}$  só é definido para uma equação  $i$  que verifique a condição de ordem, ou seja,  $k_i^* \geq M_i - 1$ , referida em (2.3.11).

Recorde-se que a condição de ordem traduz uma condição necessária para a identificabilidade da equação  $i$ , conforme (2.3.11). Como esta condição é necessária mas não é suficiente para a identificabilidade, uma equação pode verificar a condição de ordem e não ser identificada. Esta situação ocorre se para uma determinada equação se verificar a condição de ordem, mas não a condição de característica. Esta propriedade permite, então, concluir que o estimador  $2SLS$  é definido para uma equação não identificada, desde que ela verifique a condição de ordem para a identificabilidade.

P2SLS.4 Se a equação  $i$  for exactamente identificada, tem-se que  $\hat{\delta}_{i,2SLS} = \hat{\delta}_{i,ILS}$ .

P2SLS.5 Em relação à equação  $i$  na forma  $\mathbf{y}_i = \mathbf{Z}_i \delta_i + \mathbf{e}_i$ , considere-se a classe de estimadores de variáveis instrumentais definida por  $\hat{\delta}_{i,VI} = (\mathbf{P}' \mathbf{Z}_i)^{-1} \mathbf{P}' \mathbf{y}_i$ , onde as variáveis instrumentais para  $\mathbf{Z}_i$  são combinações lineares das variáveis exógenas, isto é, são da forma  $\mathbf{P} = \mathbf{X}\mathbf{A}$ . O estimador assintoticamente mais eficiente nesta classe de estimadores é o estimador  $2SLS$ .

Tal como o estimador  $2SLS$ , também o estimador  $ILS$  é um estimador de variáveis instrumentais, logo, pela última propriedade, resulta que o estimador  $2SLS$  é assintoticamente mais eficiente do que o estimador  $ILS$ .

Relembra-se que se tem vindo a trabalhar sob a hipótese de homocedasticidade, conforme se assumiu em (3.2.1). No caso de não se verificar homocedasticidade, o estimador  $2SLS$  pode ser adaptado, acrescentando-se um passo, correspondente à introdução da ponderação usual no método dos mínimos quadrados generalizados. De facto, ao admitir a heterocedasticidade do modelo, de acordo com o que se referiu em (2.2.5), tem-se que

$$\mathbf{\Omega}_i = E[\mathbf{e}_i \mathbf{e}_i'] = \text{diag} \begin{bmatrix} \sigma_{i1}^2 & \cdots & \sigma_{iT}^2 \end{bmatrix}, i = 1, \dots, M.$$

Daqui resulta que  $\mathbf{\Omega}$  não pode assumir a forma definida em (3.3.1), pois não se pode escrever com base no produto de Kronecker, pelo facto dos elementos da diagonal de cada bloco  $\mathbf{\Omega}_i$  não serem iguais entre si. Neste caso, a matriz  $\mathbf{\Omega}$  toma a forma seguinte

$$\mathbf{\Omega} = \begin{bmatrix} \mathbf{\Omega}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{\Omega}_M \end{bmatrix}.$$

A adaptação do estimador  $2SLS$  ao caso heterocedástico vai conduzir a uma versão do estimador  $2SLS$  que será designada por  $H2SLS$  (de *heteroscedastic 2SLS*), seguindo a sugestão de Greene (2003). Para obter este estimador, estima-se cada bloco  $\mathbf{\Omega}_i$ , aplicando o estimador de White, ou seja,  $\hat{\mathbf{\Omega}}_i = \mathbf{S}_i^2 = \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \left( y_{ti} - \mathbf{Z}_{ti} \hat{\boldsymbol{\delta}}_{i,2SLS} \right)^2$ ; posteriormente o estimador  $2SLS$  vai ser adaptado de modo a incorporar esta informação, de acordo com o princípio dos mínimos quadrados generalizados.

Assim, partindo do estimador  $2SLS$  na forma  $\hat{\boldsymbol{\delta}}_{i,2SLS} = \left( \hat{\mathbf{Z}}_i' \hat{\mathbf{Z}}_i \right)^{-1} \hat{\mathbf{Z}}_i' \mathbf{y}_i$ , referida por (3.3.26), vai substituir-se  $\hat{\mathbf{Z}}_i$  pela forma que será obtida a seguir:

$$\begin{aligned} \hat{\mathbf{Z}}_i &= \begin{bmatrix} \hat{\mathbf{Y}}_i & \mathbf{X}_i \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}_i & \mathbf{X}_i \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}_i & \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{X}_i \end{bmatrix} \\ &= \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \begin{bmatrix} \mathbf{Y}_i \\ \mathbf{X}_i \end{bmatrix} \\ &= \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z}_i. \end{aligned} \tag{3.3.31}$$

Donde vem que

$$\begin{aligned}\hat{\delta}_{i,2SLS} &= \left[ \mathbf{Z}'_i \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z}_i \right]^{-1} \mathbf{Z}'_i \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}_i \\ &= \left[ \mathbf{Z}'_i \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z}_i \right]^{-1} \mathbf{Z}'_i \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}_i.\end{aligned}\quad (3.3.32)$$

A partir desta forma, e procedendo à adaptação referida, obtém-se uma forma para o estimador  $H2SLS$

$$\hat{\delta}_{i,H2SLS} = \left[ \mathbf{Z}'_i \mathbf{X} (\mathbf{S}_i^2)^{-1} \mathbf{X}' \mathbf{Z}_i \right]^{-1} \mathbf{Z}'_i \mathbf{X} (\mathbf{S}_i^2)^{-1} \mathbf{X}' \mathbf{y}_i.$$

Neste caso, a matriz de covariâncias do estimador pode ser estimada por

$$\widehat{VA} \left[ \hat{\delta}_{i,H2SLS} \right] = \left[ (\mathbf{Z}'_i \mathbf{X}) (\mathbf{S}_i^2)^{-1} (\mathbf{X}' \mathbf{Z}_i) \right]^{-1}.$$

Recapitulando, o estimador  $2SLS$  obtém-se seguindo a ideia que o próprio nome sugere, na medida em que resulta da aplicação de duas fases:

Fase 1 Para cada equação  $i$ , estimar os parâmetros da forma reduzida por  $OLS$ , usando

$$\hat{\pi}_i = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}_i \text{ e obter os valores predictos } \hat{\mathbf{Y}}_i, \text{ através de } \hat{\mathbf{Y}}_i = \mathbf{X} \hat{\pi}_i.$$

Fase 2 Partindo da equação  $i$  na forma  $\mathbf{y}_i = \hat{\mathbf{Z}}_i \boldsymbol{\delta}_i + \bar{\mathbf{e}}_i$ , referida em (3.3.25), onde  $\hat{\mathbf{Z}}_i = \begin{bmatrix} \hat{\mathbf{Y}}_i & \mathbf{X}_i \end{bmatrix}$ , estimar  $\boldsymbol{\delta}_i$  por  $OLS$ , donde resulta o estimador  $\hat{\delta}_{i,2SLS}$ .

Repetindo o processo para todas as equações, podem obter-se estimativas de todos os parâmetros estruturais do modelo.

### 3.4 Estimadores de mínimos quadrados com informação completa

Seguindo esta forma de estimar os parâmetros, procede-se à estimação simultânea de todas as equações do sistema. Este tipo de procedimento permite usar a informação de todas as variáveis endógenas do sistema e entrar em consideração com a possível interacção existente entre as equações, por esta razão são designados por métodos de informação completa.

Para estimar os parâmetros estruturais de um  $SEM$ , com base nos  $MIC$ , importa obter outra forma para o modelo definido na forma estrutural por (2.2.2).

Assim, partindo da equação  $i$  na forma  $\mathbf{y}_i = \mathbf{Z}_i \boldsymbol{\delta}_i + \mathbf{e}_i$ , referida em (2.3.4), e considerando o sistema formado pelas  $M$  equações do modelo, pode escrever-se uma forma equivalente à equação matricial (2.2.2), do seguinte modo

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{Z}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}_1 \\ \vdots \\ \boldsymbol{\delta}_M \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_M \end{bmatrix}, \quad (3.4.1)$$

ou ainda,

$$\mathbf{y} = \mathbf{Z} \boldsymbol{\delta} + \mathbf{e}, \quad (3.4.2)$$

onde  $\mathbf{y} = \text{vec}(\mathbf{Y})$ ,  $\mathbf{Z} = \text{diag} \begin{bmatrix} \mathbf{Z}_1 & \cdots & \mathbf{Z}_M \end{bmatrix}$ ,  $\boldsymbol{\delta} = \begin{bmatrix} \boldsymbol{\delta}'_1 & \cdots & \boldsymbol{\delta}'_M \end{bmatrix}'$  e  $\mathbf{e} = \text{vec}(\mathbf{E})$ , com  $E(\mathbf{e}) = \mathbf{0}$ ,  $\boldsymbol{\Omega} = E(\mathbf{e}\mathbf{e}') = (\boldsymbol{\Sigma} \otimes \mathbf{I}_T)$ , por se ter admitido anteriormente a hipótese de homocedasticidade.

Para resolver a questão da existência de correlação entre os regressores,  $\mathbf{Z}$ , e os erros aleatórios,  $\mathbf{e}$ , do modelo pode aplicar-se um estimador de variáveis instrumentais, considerando como matriz de variáveis instrumentais  $\bar{\mathbf{W}}$  em relação a  $\mathbf{Z}$ , o que conduz a um estimador consistente da forma

$$\hat{\boldsymbol{\delta}}_{VI} = [\bar{\mathbf{W}}' \mathbf{Z}]^{-1} \bar{\mathbf{W}}' \mathbf{y}.$$

Por outro lado, para obter um estimador eficiente, pode aplicar-se o princípio dos mínimos quadrados generalizados, resultando em

$$\begin{aligned} \hat{\boldsymbol{\delta}}_{VI, GLS} &= [\bar{\mathbf{W}}' \boldsymbol{\Omega}^{-1} \mathbf{Z}]^{-1} \bar{\mathbf{W}}' \boldsymbol{\Omega}^{-1} \mathbf{y} \\ &= [\bar{\mathbf{W}}' (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T) \mathbf{Z}]^{-1} \bar{\mathbf{W}}' (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T) \mathbf{y}. \end{aligned} \quad (3.4.3)$$

Apresenta-se a seguir um estimador que segue esta forma de estimar os parâmetros combinada com o princípio dos mínimos quadrados.

### 3.4.1 Estimador de mínimos quadrados em três fases

Repare-se que, embora o estimador *2SLS* para a equação  $i$  use a informação proveniente de todas as variáveis exógenas do sistema, durante a primeira fase do método, ele ignora a informação contida nas variáveis endógenas que não aparecem na equação  $i$ , isto é, ignora os valores de  $\mathbf{Y}_i^*$ . Além disso, o estimador *2SLS* ignora a correlação existente

entre as diferentes equações. O estimador que se segue vai ser superior ao  $2SLS$ , pois permite corrigir estes aspectos, estimando em simultâneo todas as equações do sistema, por  $GLS$ .

Partindo do estimador de variáveis instrumentais definido em (3.3.7) e definindo as variáveis instrumentais

$$\bar{\mathbf{W}} = \hat{\mathbf{Z}} = \begin{bmatrix} \hat{\mathbf{Z}}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{Z}}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \hat{\mathbf{Z}}_M \end{bmatrix},$$

onde  $\hat{\mathbf{Z}}_i = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_i$ ,  $i = 1, \dots, M$ , de acordo com a expressão obtida em (3.3.31), resulta um estimador da forma

$$\hat{\boldsymbol{\delta}}_{VI} = [\hat{\mathbf{Z}}'\mathbf{Z}]^{-1}\hat{\mathbf{Z}}'\mathbf{y}.$$

Note-se que este estimador corresponde ao estimador  $2SLS$  calculado para todas as equações do modelo. Aplicando de seguida o método  $GLS$ , obtém-se um estimador definido por

$$\hat{\boldsymbol{\delta}}_{3SLS} = [\hat{\mathbf{Z}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T)\mathbf{Z}]^{-1}\hat{\mathbf{Z}}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T)\mathbf{y}. \quad (3.4.4)$$

Repare-se que esta forma corresponde ao estimador referido em (3.4.3) quando se consideram  $\hat{\mathbf{Z}}$  como variáveis instrumentais.

Para obter a forma do estimador  $3SLS$  para  $\boldsymbol{\delta}$ , considere-se o modelo na forma (3.4.1) e multiplique-se à esquerda por  $\mathbf{X}'$ , donde resulta que

$$\begin{bmatrix} \mathbf{X}'\mathbf{y}_1 \\ \vdots \\ \mathbf{X}'\mathbf{y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{Z}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{X}'\mathbf{Z}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}_1 \\ \vdots \\ \boldsymbol{\delta}_M \end{bmatrix} + \begin{bmatrix} \mathbf{X}'\mathbf{e}_1 \\ \vdots \\ \mathbf{X}'\mathbf{e}_M \end{bmatrix}, \quad (3.4.5)$$

ou ainda, de forma compacta,

$$(\mathbf{I}_M \otimes \mathbf{X}')\mathbf{y} = (\mathbf{I}_M \otimes \mathbf{X}')\mathbf{Z}\boldsymbol{\delta} + (\mathbf{I}_M \otimes \mathbf{X}')\mathbf{e}. \quad (3.4.6)$$

Na expressão anterior, as dimensões das matrizes são as seguintes

$$\underbrace{\underbrace{(\mathbf{I}_M \otimes \mathbf{X}')}_{(Mk \times MT)} \underbrace{\mathbf{y}}_{(MT \times 1)}}_{(Mk \times 1)} = \underbrace{\underbrace{(\mathbf{I}_M \otimes \mathbf{X}')}_{(Mk \times MT)} \underbrace{\mathbf{Z}}_{\left(MT \times \sum_{i=1}^M (M_i + k_i - 1)\right)} \underbrace{\boldsymbol{\delta}}_{\left(\sum_{i=1}^M (M_i + k_i - 1) \times 1\right)}}_{(Mk \times 1)} + \underbrace{\underbrace{(\mathbf{I}_M \otimes \mathbf{X}')}_{(Mk \times MT)} \underbrace{\mathbf{e}}_{(MT \times 1)}}_{(Mk \times 1)}.$$



Para obter a expressão do estimador  $GLS$ , é necessário entrar com a matriz de covariâncias dos erros do modelo. Desde já se verifica que ela pode ser expressa numa forma especialmente conveniente; atendendo a (3.2.3), vem que

$$\begin{aligned} V[(\mathbf{I}_M \otimes \mathbf{X}') \mathbf{e}] &= E[(\mathbf{I}_M \otimes \mathbf{X}') \mathbf{e} \mathbf{e}' (\mathbf{I}_M \otimes \mathbf{X})] \\ &= (\mathbf{I}_M \otimes \mathbf{X}') E[\mathbf{e} \mathbf{e}'] (\mathbf{I}_M \otimes \mathbf{X}) \\ &= (\mathbf{I}_M \otimes \mathbf{X}') (\boldsymbol{\Sigma} \otimes \mathbf{I}_T) (\mathbf{I}_M \otimes \mathbf{X}), \\ &= (\boldsymbol{\Sigma} \otimes \mathbf{X}' \mathbf{X}). \end{aligned} \quad (3.4.7)$$

Deste modo, uma vez que o estimador de  $GLS$  de um parâmetro  $\beta$  no modelo  $\mathbf{y} = \beta \mathbf{X} + \varepsilon$ , é da forma genérica  $\hat{\beta} = (\mathbf{X}' \mathbf{S}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{S}^{-1} \mathbf{y}$ , onde  $\mathbf{S} = \widehat{\mathbf{V}}[\varepsilon]$ , da equação (3.4.6) resulta que o estimador  $GLS$  para  $\delta$  é

$$\begin{aligned} \hat{\delta}_{GLS} &= \left\{ [(\mathbf{I}_M \otimes \mathbf{X}') \mathbf{Z}]' [\boldsymbol{\Sigma} \otimes (\mathbf{X}' \mathbf{X})]^{-1} [(\mathbf{I}_M \otimes \mathbf{X}') \mathbf{Z}] \right\}^{-1} \\ &\quad \times [(\mathbf{I}_M \otimes \mathbf{X}') \mathbf{Z}]' [\boldsymbol{\Sigma} \otimes (\mathbf{X}' \mathbf{X})]^{-1} (\mathbf{I}_M \otimes \mathbf{X}') \mathbf{y} \\ &= \left\{ \mathbf{Z}' [\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'] \mathbf{Z} \right\}^{-1} \mathbf{Z}' [\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'] \mathbf{y}. \end{aligned} \quad (3.4.8)$$

Este estimador pode ser obtido de outro modo, como se vai ver de seguida. Tendo em conta que  $\mathbf{X}' \mathbf{X}$  é uma matriz semidefinida positiva, existe sempre uma matriz  $\mathbf{P}_{(k \times k)}$ , tal que  $\mathbf{P} (\mathbf{X}' \mathbf{X}) \mathbf{P}' = \mathbf{I}_k$ . Assim, obtém-se outra forma para (3.4.5),

$$\begin{bmatrix} \mathbf{P} \mathbf{X}' \mathbf{y}_1 \\ \vdots \\ \mathbf{P} \mathbf{X}' \mathbf{y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{P} \mathbf{X}' \mathbf{Z}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{P} \mathbf{X}' \mathbf{Z}_M \end{bmatrix} \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_M \end{bmatrix} + \begin{bmatrix} \mathbf{P} \mathbf{X}' \mathbf{e}_1 \\ \vdots \\ \mathbf{P} \mathbf{X}' \mathbf{e}_M \end{bmatrix},$$

ou de forma compacta,

$$(\mathbf{I}_M \otimes \mathbf{P} \mathbf{X}') \mathbf{y} = (\mathbf{I}_M \otimes \mathbf{P} \mathbf{X}') \mathbf{Z} \delta + (\mathbf{I}_M \otimes \mathbf{P} \mathbf{X}') \mathbf{e}. \quad (3.4.9)$$

Reescrevendo uma única equação para o modelo, ela será da forma,

$$\check{\mathbf{y}} = \check{\mathbf{Z}} \delta + \check{\mathbf{e}}, \quad (3.4.10)$$

onde  $\check{\mathbf{y}} = (\mathbf{I}_M \otimes \mathbf{P} \mathbf{X}') \mathbf{y}$ ,  $\check{\mathbf{Z}} = (\mathbf{I}_M \otimes \mathbf{P} \mathbf{X}') \mathbf{Z}$  e  $\check{\mathbf{e}} = (\mathbf{I}_M \otimes \mathbf{P} \mathbf{X}') \mathbf{e}$ . Neste caso, a matriz de covariâncias dos erros é da forma

$$\begin{aligned} V[\check{\mathbf{e}}] &= V[(\mathbf{I}_M \otimes \mathbf{P} \mathbf{X}') \mathbf{e}] \\ &= [\boldsymbol{\Sigma} \otimes \mathbf{P} (\mathbf{X}' \mathbf{X}) \mathbf{P}'] \\ &= (\boldsymbol{\Sigma} \otimes \mathbf{I}_k), \end{aligned}$$

fazendo um desenvolvimento análogo ao seguido em (3.4.7).

Aplicando o método *GLS* a (3.4.10), vem que

$$\begin{aligned}\hat{\delta}_{GLS} &= \left[ \check{\mathbf{Z}}' (\boldsymbol{\Sigma} \otimes \mathbf{I}_k)^{-1} \check{\mathbf{Z}} \right]^{-1} \check{\mathbf{Z}}' (\boldsymbol{\Sigma} \otimes \mathbf{I}_k)^{-1} \check{\mathbf{y}} \\ &= \left\{ [(\mathbf{I}_M \otimes \mathbf{P}\mathbf{X}') \mathbf{Z}]' (\boldsymbol{\Sigma} \otimes \mathbf{I}_k)^{-1} [(\mathbf{I}_M \otimes \mathbf{P}\mathbf{X}') \mathbf{Z}] \right\}^{-1} \\ &\quad \times [(\mathbf{I}_M \otimes \mathbf{P}\mathbf{X}') \mathbf{Z}]' (\boldsymbol{\Sigma} \otimes \mathbf{I}_k)^{-1} (\mathbf{I}_M \otimes \mathbf{P}\mathbf{X}') \mathbf{y} \\ &= [\mathbf{Z}' (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X}\mathbf{P}'\mathbf{P}\mathbf{X}') \mathbf{Z}]^{-1} \mathbf{Z}' (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X}\mathbf{P}'\mathbf{P}\mathbf{X}') \mathbf{y}.\end{aligned}\quad (3.4.11)$$

Mas, como a partir de  $\mathbf{P}(\mathbf{X}'\mathbf{X})\mathbf{P}' = \mathbf{I}_k$  se obtém que  $\mathbf{P}'\mathbf{P} = (\mathbf{X}'\mathbf{X})^{-1}$ , a expressão (3.4.11) simplifica-se em

$$\hat{\delta}_{GLS} = \left\{ \mathbf{Z}' \left[ \boldsymbol{\Sigma}^{-1} \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right] \mathbf{Z} \right\}^{-1} \mathbf{Z}' \left[ \boldsymbol{\Sigma}^{-1} \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right] \mathbf{y}.$$

Repare-se que a forma do estimador  $\hat{\delta}_{GLS}$  agora encontrada coincide com a obtida pelo processo anterior. No entanto, esta forma depende de  $\boldsymbol{\Sigma}$ . No caso desta matriz ser desconhecida, pode usar-se um estimador consistente que será representado por  $\hat{\boldsymbol{\Sigma}} = [\hat{\sigma}_{ij}]_{i,j=1,\dots,M}$ , onde  $\hat{\sigma}_{ij}$  são as covariâncias amostrais dos resíduos, calculados com base nas estimativas  $\hat{\delta}_{i,2SLS}$  (resultantes da aplicação do estimador *2SLS* a cada equação), isto é,

$$\hat{\sigma}_{ij} = \frac{1}{T} \left[ \left( \mathbf{y}_i - \mathbf{Z}_i \hat{\delta}_{i,2SLS} \right)' \left( \mathbf{y}_j - \mathbf{Z}_j \hat{\delta}_{j,2SLS} \right) \right]. \quad (3.4.12)$$

Obtém-se finalmente a forma do estimador *3SLS* para  $\boldsymbol{\delta}$ ,

$$\hat{\delta}_{3SLS} = \left[ \mathbf{Z}' \left( \hat{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right) \mathbf{Z} \right]^{-1} \mathbf{Z}' \left( \hat{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right) \mathbf{y}. \quad (3.4.13)$$

O estimador  $\hat{\delta}_{3SLS}$  goza das propriedades que se apresentam de seguida. As demonstrações das duas primeiras propriedades podem ser vistas, respectivamente, em Greene (2003) e Judge *et al.* (1988) e as restantes em Schmidt (1976).

P3SLS.1 O estimador  $\hat{\delta}_{3SLS}$  é um estimador de variáveis instrumentais.

P3SLS.2 O estimador  $\hat{\delta}_{3SLS}$  é um estimador consistente para  $\boldsymbol{\delta}$  e tem-se que

$$\sqrt{T} \left( \hat{\delta}_{3SLS} - \boldsymbol{\delta} \right) \overset{\bullet}{\sim} N \left( \mathbf{0}, \left[ \boldsymbol{\Sigma}_{\mathbf{Z}\mathbf{X}} (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1}) \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Z}} \right]^{-1} \right).$$

Um estimador consistente da matriz de covariâncias é

$$\widehat{VA} \left( \hat{\delta}_{3SLS} \right) = \left\{ \frac{1}{T} \mathbf{Z}' \left[ \hat{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right] \mathbf{Z} \right\}^{-1}, \text{ onde } \hat{\boldsymbol{\Sigma}} \text{ tem elementos } \hat{\sigma}_{ij} \text{ definidos em (3.4.12).}$$

P3SLS.3 O estimador  $\hat{\delta}_{3SLS}$  é assintoticamente mais eficiente do que o estimador  $\hat{\delta}_{2SLS}$ .

P3SLS.4 Se todas as equações forem exactamente identificadas, tem-se que  $\hat{\delta}_{3SLS} = \hat{\delta}_{2SLS}$ , onde

$$\hat{\delta}_{2SLS} = \begin{bmatrix} \hat{\delta}'_{1,2SLS} & \cdots & \hat{\delta}'_{M,2SLS} \end{bmatrix}' \quad (3.4.14)$$

é o vector de estimativas obtidas aplicando o estimador  $2SLS$  às  $M$  equações estruturais do modelo.

P3SLS.5 Se alguma equação do modelo não verificar a condição de ordem para a identificabilidade (**Proposição 2.3.3**), o estimador  $3SLS$  não existe.

No entanto se tal acontecer, pode abandonar-se alguma equação que seja subidentificada e aplicar-se o estimador  $3SLS$  às restantes equações do modelo.

Resumindo, o estimador  $3SLS$  obtém-se aplicando sucessivamente as três fases seguintes:

Fase 1 Para cada equação, estimar os parâmetros  $\pi_i$  da forma reduzida por  $OLS$  usando

$$\hat{\pi}_i = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}_i \text{ e calcular os valores predictos } \hat{\mathbf{Y}}_i = \mathbf{X}\hat{\pi}_i \text{ e } \hat{\mathbf{Z}}_i = \begin{bmatrix} \hat{\mathbf{Y}}_i & \mathbf{X}_i \end{bmatrix}.$$

Fase 2 Calcular  $\hat{\delta}_{i,2SLS}$  para cada uma das equações ( $i = 1, \dots, M$ ) e obter uma estimativa da matriz de covariâncias dos erros  $\hat{\Sigma} = [\hat{\sigma}_{ij}]$ , com  $\hat{\sigma}_{ij}$  definido em (3.4.12).

Fase 3 Usando  $\hat{\Sigma}$  obtida na fase anterior, estimar  $\delta$  usando  $\hat{\delta}_{3SLS}$  definido em (3.4.13).

A partir da estimativa  $\hat{\delta}_{3SLS}$ , pode obter-se uma versão do estimador  $3SLS$  ligeiramente modificada, vulgarmente designada por  $\hat{\delta}_{I3SLS}$  (de *iterative three-stage least squares*). De acordo com esta modificação, os estimadores são calculados pelos seguintes procedimentos: obtêm-se valores dos resíduos  $3SLS$ , os quais são usados no cálculo de uma nova estimativa da matriz de covariâncias,  $\hat{\Sigma}$ ; com esta matriz calcula-se uma nova estimativa  $\hat{\delta}_{3SLS}$ , aplicando (3.4.13). A estimativa resultante é  $\hat{\delta}_{I3SLS}$ . Embora o processo possa ser repetido até não haver alterações significativas nas estimativas dos coeficientes estruturais, Kmenta (1997) defende que o estimador  $\hat{\delta}_{I3SLS}$  não apresenta melhores propriedades assintóticas do que o estimador  $\hat{\delta}_{3SLS}$ .

### 3.5 Comparação das matrizes de covariâncias assintóticas dos estimadores *2SLS* e *3SLS*

Para comparar o estimador  $\hat{\delta}_{3SLS}$  com o estimador  $\hat{\delta}_{2SLS}$ , pode-se efectuar o estudo da eficiência relativa assintótica, que é estabelecido pela comparação das matrizes de covariâncias assintóticas dos estimadores. De acordo com Staudte e Sheather (1990), esta comparação pode ser feita de acordo com um dos três critérios que se apresentam de seguida.

Critério.1 O estimador  $T_1$  é mais eficiente do que o estimador  $T_2$  se

$$V[T_1] \leq V[T_2]$$

e

$$V[T_1] \leq V[T_2] \Leftrightarrow V[T_2] - V[T_1] \geq 0,$$

o que traduz que a matriz  $V[T_2] - V[T_1]$  é semidefinida positiva.

Critério.2 (variância generalizada) O estimador  $T_1$  é mais eficiente do que o estimador  $T_2$  se

$$\det V[T_1] \leq \det V[T_2].$$

Critério.3 (variação total) O estimador  $T_1$  é mais eficiente do que o estimador  $T_2$  se

$$tr(V[T_1]) \leq tr(V[T_2]),$$

onde  $tr(\mathbf{A})$  representa o traço da matriz  $\mathbf{A}$ .

Para comparar os estimadores *3SLS* e *2SLS*, verificou-se que o estimador  $\hat{\delta}_{3SLS}$  é assintoticamente mais eficiente do que o estimador  $\hat{\delta}_{2SLS}$  (a demonstração pode ser vista em Judge *et al.* (1988)). Para isso, aplicando o Critério.1, mostrou-se que a matriz  $VA[\hat{\delta}_{2SLS}] - VA[\hat{\delta}_{3SLS}]$  é semidefinida positiva, donde resultou que,

$$VA[\hat{\delta}_{2SLS}] \geq VA[\hat{\delta}_{3SLS}].$$

De acordo com Zellner e Theil (1962), há apenas dois casos em que não se verifica a superioridade de  $\hat{\delta}_{3SLS}$  em relação a  $\hat{\delta}_{2SLS}$ , ou seja, em que as correspondentes matrizes de covariâncias são assintoticamente equivalentes, o que significa que os processos conduzem à mesma eficiência assintótica. Essas situações ocorrem quando não há correlação entre as equações, isto é, quando  $\Sigma$  é uma matriz diagonal, ou quando todas as equações do sistema são exactamente identificadas. Posteriormente, Srivastava e Tiwari (1978) estabelecem condições mais gerais para que os estimadores de  $2SLS$  e  $3SLS$  sejam igualmente eficientes e Kapteyn e Fiebig (1981) obtêm uma condição necessária e suficiente para a equivalência destes estimadores.

### 3.6 Utilização do Método dos Momentos Generalizado em *SEM*

O princípio subjacente ao Método dos Momentos consiste em obter estimadores dos parâmetros como soluções de um sistema de equações, resultantes de igualar os momentos populacionais aos momentos amostrais correspondentes (condições dos momentos), tendo em conta a convergência dos segundos para os primeiros, em condições de regularidade. A aplicação do Método dos Momentos exige que se disponha de uma amostra com componentes independentes e idênticamente distribuídas e que o sistema formado pelas condições dos momentos tenha igual número de equações e de parâmetros a estimar. Neste caso, o sistema é possível e determinado, podendo obter-se formas explícitas para os estimadores.

O Método dos Momentos Generalizado (*GMM*) generaliza o Método dos Momentos ao estimar, pelo Método dos Momentos, a esperança de funções de parâmetros que traduzem condições do modelo em estudo.

O *GMM* foi desenvolvido por Hansen (1982) e apresenta como vantagem o facto de requerer apenas a verificação de algumas condições envolvendo os momentos, não impondo a homocedasticidade dos erros do modelo e podendo aplicar-se quando o número de parâmetros desconhecidos é menor do que o número de condições dos momentos.

A estimação por *GMM* encontra-se discutida e desenvolvida com detalhe em livros publicados por diversos autores, como por exemplo, Hamilton (1994), Cochrane (2001),

Arellano (2003), Hall (2005) e Singleton (2006). Ghysels e Hall (2002) elaboraram um apanhado dos principais desenvolvimentos da estimação por *GMM*. De entre muitos artigos publicados sobre a teoria estatística subjacente ao *GMM* e sobre o impacto deste método na teoria econométrica, importa destacar Hall (1993 e 2001), Newey e McFadden (1994) e Imbens (2002). A importância do *GMM* em Econometria pode ser avaliada através de uma série de artigos que reflectem as principais áreas de investigação activa. O maior impacto deve verificar-se na análise econométrica aplicada nas áreas de Macroeconomia e de Finanças: Hansen e West (2002) avaliaram a influência do *GMM* na Macroeconomia e Jaganathan *et al.* (2002) fizeram um levantamento do seu impacto nas Finanças. Mais no âmbito do estudo estatístico do estimador *GMM*, Imbens e Spady (2002) desenvolveram intervalos de confiança para parâmetros caracterizados por restrições de momentos, com base no método da verosimilhança empírica. Por outro lado, tendo em consideração que as propriedades do estimador *GMM* dependem da escolha das variáveis instrumentais, Donald *et al.* (2009) propuseram e implementaram um erro quadrático médio assintótico, baseado em critérios de selecção de instrumentos. Importa ainda referir que a estimação por *GMM* continua a ser um tema muito estudado, quer em termos teóricos, quer em aplicações, a avaliar pela quantidade de trabalhos publicados recentemente sobre este assunto.

Para ajudar a clarificar a ideia da estimação por *GMM*, veja-se o seguinte exemplo publicado em Hamilton (1994).

*Exemplo 3.6.1.* Considere-se uma variável aleatória  $Y_t$  com distribuição t de Student com  $\nu$  graus de liberdade. Suponha-se que se dispõe duma amostra de observações independentes e identicamente distribuídas (*i.i.d.*) com dimensão  $T$ ,  $(y_1, \dots, y_T)$ , e que se pretende estimar  $\nu$ .

Recorde-se que, para a distribuição t-Student com  $\nu$  graus de liberdade e  $\nu > 2$ , se tem que

$$E[Y_t] = 0 \text{ e } V[Y_t] = E[Y_t^2] = \mu_2 = \frac{\nu}{\nu - 2}.$$

Quando  $T$  tende para infinito, o momento amostral de segunda ordem,

$$M_2 = \frac{1}{T} \sum_{t=1}^T y_t^2,$$

converge em probabilidade para  $\mu_2$ , ou seja,

$$p \lim_{T \rightarrow \infty} M_2 = \mu_2,$$

e  $\hat{\mu}_2 = M_2$  é um estimador consistente para  $\mu_2$ .

Assim, estabelecendo a equação

$$\frac{\nu}{\nu - 2} = M_2,$$

obtém-se o estimador

$$\hat{\nu}_{MM} = \frac{2M_2}{M_2 - 1}.$$

Assim,  $\hat{\nu}_{MM}$  é o estimador de  $\nu$  obtido pelo Método dos Momentos. Note-se que para obter o estimador  $\hat{\nu}_{MM}$  se usou apenas um momento amostral - neste caso foi  $M_2$ . Suponha-se que havia interesse em outro momento, por exemplo, o momento amostral de ordem quatro. Nesse caso, o momento populacional seria

$$E[Y_t^4] = \mu_4 = \frac{3\nu^2}{(\nu - 2)(\nu - 4)},$$

para  $\nu > 4$ , enquanto que o momento amostral da mesma ordem é definido por

$$M_4 = \frac{1}{T} \sum_{t=1}^T y_t^4.$$

Tal como no caso dos momentos de segunda ordem, também  $M_4$  é um estimador consistente de  $\mu_4$ , podendo também ser usado para obter um estimador para  $\nu$ . Assim, fica estabelecido um sistema de duas equações dependentes de um só parâmetro,

$$M_2 = \frac{\nu}{\nu - 2} \text{ e } M_4 = \frac{3\nu^2}{(\nu - 2)(\nu - 4)},$$

e o método dos momentos generalizado vai ser usado. Embora não se possa obter uma expressão única para  $\nu$  que verifique ambas as condições, pode escolher-se para  $\nu$  o valor que melhor aproxima as soluções de ambas as equações. Isto corresponde a atribuir a  $\nu$  o valor da solução do problema de minimização de

$$\mathbf{Q}(\nu) = \mathbf{g}'\mathbf{W}\mathbf{g}, \quad (3.6.1)$$

onde

$$\mathbf{g} = \begin{bmatrix} M_2 - \frac{\nu}{\nu - 2} \\ M_4 - \frac{3\nu^2}{(\nu - 2)(\nu - 4)} \end{bmatrix}$$

e  $\mathbf{W}$  é uma matriz de ponderação  $(2 \times 2)$ , simétrica e definida positiva. Quanto maior for o elemento de  $\mathbf{W}$ , maior é a importância dada à correspondente condição do momento.  $\diamond$

Um estimador obtido como solução de um problema do tipo (3.6.1) é designado por estimador de distância mínima, por vários autores. Hansen (1982) apresenta uma caracterização mais geral desta aproximação, deduz alguns resultados assintóticos para o estimador obtido por esta via e designa este processo de estimação por Método dos Momentos Generalizado. Na formulação deste método seguiu-se uma importante contribuição de Hamilton (1994), que se passa a resumir.

Considere-se  $\mathbf{w}_t$  um vector aleatório  $h$ -dimensional, de observações no momento  $t$ ,  $\boldsymbol{\theta}$  um vector de parâmetros desconhecidos, de dimensão  $(a \times 1)$  e a função vectorial

$$\begin{aligned} \mathbf{h} : R^a \times R^h &\longrightarrow R^r \\ (\boldsymbol{\theta}, \mathbf{w}_t) &\longrightarrow \mathbf{h}(\boldsymbol{\theta}, \mathbf{w}_t). \end{aligned}$$

Suponha-se que  $\boldsymbol{\theta}_0$  representa o verdadeiro valor de  $\boldsymbol{\theta}$ , que é desconhecido, e que verifica a condição

$$E[\mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_t)] = \mathbf{0}. \quad (3.6.2)$$

A equação vectorial (3.6.2) corresponde a um sistema de  $r$  equações, correspondentes às condições dos momentos e que são habitualmente designadas por condições de ortogonalidade.

Considere-se que  $\mathbb{Y}_T = (\mathbf{w}'_T, \mathbf{w}'_{T-1}, \dots, \mathbf{w}'_1)'$  é um vector de dimensão  $(Th \times 1)$ , formado por todas as observações de uma amostra de dimensão  $T$  e represente-se a média amostral de  $\mathbf{h}(\boldsymbol{\theta}, \mathbf{w}_t)$  por  $\mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T)$ , ou seja,

$$\mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T) = \frac{1}{T} \sum_{t=1}^T \mathbf{h}(\boldsymbol{\theta}, \mathbf{w}_t). \quad (3.6.3)$$

A ideia base do *GMM* consiste em escolher para  $\boldsymbol{\theta}$  o valor que faz com que o momento amostral  $\mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T)$  esteja o mais próximo possível do primeiro momento populacional  $E[\mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_t)]$ , que é nulo. Isto traduz que  $\widehat{\boldsymbol{\theta}}_T$ , o estimador *GMM* para  $\boldsymbol{\theta}$ , é obtido como solução do problema de minimização



$$\min_{\boldsymbol{\theta}} \mathbf{Q}(\boldsymbol{\theta}, \mathbb{Y}_T) = \min_{\boldsymbol{\theta}} \{ [\mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T)]' \mathbf{W}_T [\mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T)] \}, \quad (3.6.4)$$

onde  $\{\mathbf{W}_T\}$  é uma sucessão de matrizes definidas positivas, de dimensão  $(r \times r)$ , que podem ser função dos dados contidos em  $\mathbb{Y}_T$ .

A solução do problema (3.6.4) pode ser encontrada com menos esforço computacional, resolvendo o problema

$$\mathbf{g}(\widehat{\boldsymbol{\theta}}_T, \mathbb{Y}_T) = \mathbf{0}, \quad (3.6.5)$$

uma vez que as soluções desta equação verificam as condições de estacionaridade do problema de minimização.

Apresentam-se alguns comentários sobre a discussão do sistema. Quando o número de parâmetros desconhecidos coincide com o número de condições de ortogonalidade, o estimador *GMM* para  $\boldsymbol{\theta}$  é a solução única do sistema de  $r$  equações (3.6.5). Quando o número de parâmetros desconhecidos é inferior ao número de condições de ortogonalidade, o sistema (3.6.5) não tem uma solução única. A distância da componente  $i$  de  $\mathbf{g}(\widehat{\boldsymbol{\theta}}_T, \mathbb{Y}_T)$  à origem é tanto menor quanto maior for o peso da  $i$ -ésima componente da condição de ortogonalidade, o qual é especificado pelo elemento  $(i, i)$  da matriz de pesos  $\mathbf{W}_T$ .

Desde já se pode verificar que a solução de (3.6.4) é um estimador consistente para  $\boldsymbol{\theta}$ , sob condições gerais de estacionaridade, de continuidade e condições de momentos (estas condições podem ser vistas em Hansen, 1982). De facto, para qualquer valor de  $\boldsymbol{\theta}$ , o vector  $\mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T)$  corresponde à média amostral de  $T$  realizações do vector  $\mathbf{h}(\boldsymbol{\theta}, \mathbf{w}_t)$ , de acordo com (3.6.3). Para  $\mathbf{h}$  contínua e  $\mathbf{w}_t$  estritamente estacionário, pode aplicar-se a Lei dos Grandes Números, pelo que se verifica

$$p \lim_{T \rightarrow \infty} \mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T) = E[\mathbf{h}(\boldsymbol{\theta}, \mathbf{w}_t)].$$

A esperança  $E[\mathbf{h}(\boldsymbol{\theta}, \mathbf{w}_t)]$  depende do valor de  $\boldsymbol{\theta}$  e da distribuição de probabilidade de  $\mathbf{w}_t$ .

O desenvolvimento que se apresenta a seguir, explica como se determina a matriz de pesos óptima.

Quando se considera o verdadeiro valor do parâmetro,  $\boldsymbol{\theta}_0$ , o processo  $\{\mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_t)\}$

é estritamente estacionário com média zero e  $\nu$ -ésima matriz de covariâncias

$$\mathbf{\Gamma}_\nu = E [\mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_t) (\mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_{t-\nu}))'] .$$

Admita-se que a série  $\sum_{\nu=-\infty}^{\infty} \mathbf{\Gamma}_\nu$  é absolutamente convergente com soma  $\mathbf{S}$ , isto é, que  $\mathbf{S} = \sum_{\nu=-\infty}^{\infty} \mathbf{\Gamma}_\nu$ . Como  $\mathbf{g}(\boldsymbol{\theta}_0, \mathbb{Y}_T)$  é a média amostral de  $\mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_t)$ , então, pelas propriedades de um processo estacionário, tem-se que  $\mathbf{S}$  é a matriz de covariâncias assintótica de  $\mathbf{g}(\boldsymbol{\theta}_0, \mathbb{Y}_T)$ ; ou seja,

$$\mathbf{S} = p \lim_{T \rightarrow \infty} TE [(\mathbf{g}(\boldsymbol{\theta}_0, \mathbb{Y}_T)) (\mathbf{g}(\boldsymbol{\theta}_0, \mathbb{Y}_T))'] .$$

A matriz de pesos óptima em (3.6.4) vai ser a inversa desta matriz de covariâncias assintótica, ou seja,  $\mathbf{W}_T = \mathbf{S}^{-1}$ . Doutro modo, pode dizer-se que a matriz de covariâncias assintótica do estimador *GMM* é mínima quando  $\widehat{\boldsymbol{\theta}}_T$  é obtido como solução do problema  $\min_{\boldsymbol{\theta}} \mathbf{Q}(\boldsymbol{\theta}, \mathbb{Y}_T)$ , onde  $\mathbf{Q}(\boldsymbol{\theta}, \mathbb{Y}_T) = [\mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T)]' \mathbf{S}^{-1} [\mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T)]$ .

Do ponto de vista intuitivo, considere-se um modelo linear simples em relação ao qual se dispõe de  $r$  observações  $(y_1, \dots, y_r)$ , com uma média populacional  $(\mu_1, \dots, \mu_r)$  e, portanto, diferente para cada observação. Deste modo,  $y_1$  pode ser interpretada como a média amostral de uma amostra de  $T_1$  observações,  $y_2$  a média amostral de uma segunda amostra de  $T_2$  observações, e assim sucessivamente. Não havendo restrições, as estimativas seriam  $\hat{\mu}_i = y_i$ , para  $i = 1, \dots, r$ . No caso de haver restrições lineares entre os  $\mu_i$ , os melhores estimadores, são definidos por funções lineares dos  $y_i$  e podem ser obtidos por *GLS*. Recorde-se que o estimador *GLS* para  $\boldsymbol{\mu}$  é a solução do problema

$$\min_{\boldsymbol{\mu}} (\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Omega}^{-1} (\mathbf{y} - \boldsymbol{\mu}) , \quad (3.6.6)$$

onde  $\mathbf{y} = (y_1, \dots, y_r)'$ ,  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_r)'$  e  $\boldsymbol{\Omega} = E [(\mathbf{y} - \boldsymbol{\mu}) (\mathbf{y} - \boldsymbol{\mu})']$  é a matriz de covariâncias de  $(\mathbf{y} - \boldsymbol{\mu})$ . A matriz de pesos óptima usada na função a minimizar em (3.6.6) é  $\boldsymbol{\Omega}^{-1}$ . Do mesmo modo que a matriz  $\boldsymbol{\Omega}$  em (3.6.6) é a matriz de covariâncias de  $(\mathbf{y} - \boldsymbol{\mu})$ , também a matriz  $\mathbf{S}$ , anteriormente referida, é a matriz de covariâncias assintótica de  $[\sqrt{T} \mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T)]$ . Como o processo é não correlacionado em relação a  $t$ , um estimador consistente para a matriz  $\mathbf{S}$  ficaria definido por

$$\mathbf{S}_T^* = \frac{1}{T} \sum_{t=1}^T [\mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_t)] [\mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_t)]' ,$$

caso  $\boldsymbol{\theta}_0$  fosse conhecido. Estimando  $\boldsymbol{\theta}_0$  através de  $\widehat{\boldsymbol{\theta}}_T$  decorre que

$$\hat{\mathbf{S}}_T = \frac{1}{T} \sum_{t=1}^T \left[ \mathbf{h}(\widehat{\boldsymbol{\theta}}_T, \mathbf{w}_t) \right] \left[ \mathbf{h}(\widehat{\boldsymbol{\theta}}_T, \mathbf{w}_t) \right]', \quad (3.6.7)$$

onde

$$p \lim_{T \rightarrow \infty} \hat{\mathbf{S}}_T = \mathbf{S}, \quad (3.6.8)$$

para algum estimador consistente  $\widehat{\boldsymbol{\theta}}_T$  de  $\boldsymbol{\theta}$ .

Repare-se que de acordo com o raciocínio acabado de descrever, o objectivo inicial era estimar  $\boldsymbol{\theta}$ , sendo para isso necessário estimar  $\mathbf{S}$ ; mas a estimação de  $\mathbf{S}$  necessita da estimação prévia de  $\boldsymbol{\theta}$ . Assim, para resolver esta questão e possibilitar a aplicação do estimador *GMM*, é necessário dispor de uma estimativa inicial de  $\boldsymbol{\theta}$ ,  $\boldsymbol{\theta}_0$ , a qual pode ser obtida resolvendo o problema (3.6.4), considerando a matriz de pesos  $\mathbf{W}_T = \mathbf{I}_T$ . Esta estimativa é depois usada em (3.6.7) para obter uma estimativa inicial de  $\hat{\mathbf{S}}_T$ , designada por  $\hat{\mathbf{S}}_T^0$ . De seguida, procede-se de novo à resolução do problema (3.6.4), agora com matriz de pesos  $\mathbf{W}_T = \left( \hat{\mathbf{S}}_T^0 \right)^{-1}$ , para obter uma nova estimativa para  $\boldsymbol{\theta}$ , que é a estimativa obtida por *GMM*,  $\widehat{\boldsymbol{\theta}}_T$ . Este processo pode ser repetido, fazendo mais iterações, embora o estimador obtido com base numa iteração seja assintoticamente equivalente ao que se obtém com um elevado número de iterações.

Para estabelecer a distribuição normal assintótica do estimador *GMM*, considere-se  $\widehat{\boldsymbol{\theta}}_T$  definido como solução do problema de

$$\min_{\boldsymbol{\theta}} [\mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T)]' \hat{\mathbf{S}}_T^{-1} [\mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T)],$$

onde  $\hat{\mathbf{S}}_T^{-1}$  é constante em relação a  $\boldsymbol{\theta}$  e tal que  $p \lim_{T \rightarrow \infty} \hat{\mathbf{S}}_T = \mathbf{S}$ . O estimador *GMM* pode ser obtido resolvendo o sistema

$$\left\{ \frac{\partial \mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T)}{\partial \boldsymbol{\theta}'} \right\}' \hat{\mathbf{S}}_T^{-1} \mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T) = \mathbf{0}. \quad (3.6.9)$$

Admitindo que  $E \{ [\mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_t)] [\mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_{t-\nu})]' \} = \boldsymbol{\Gamma}_\nu$  não depende de  $t$ , como  $\mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T)$  é a média amostral de um processo cuja esperança é nula,  $\mathbf{g}$  deve satisfazer o Teorema do Limite Central, e

$$\sqrt{T} \mathbf{g}(\boldsymbol{\theta}_0, \mathbb{Y}_T) \overset{\bullet}{\sim} N(\mathbf{0}, \mathbf{S}). \quad (3.6.10)$$

Com base em (3.6.10), surge a proposição seguinte, que se encontra demonstrada em Hamilton (1994).

**Proposição 3.6.1.** Seja  $\mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T)$  diferenciável em  $\boldsymbol{\theta}$ , para todo o  $\mathbb{Y}_T$ , e seja  $\widehat{\boldsymbol{\theta}}_T$  o estimador *GMM* satisfazendo (3.6.9), com  $r \geq a$ . Seja  $(\hat{\mathbf{S}}_T)$  uma sucessão de matrizes  $(r \times r)$  definidas positivas e tais que  $p \lim_{T \rightarrow \infty} \hat{\mathbf{S}}_T = \mathbf{S}$ , onde  $\mathbf{S}$  é uma matriz definida positiva. Suponha-se ainda que se verificam as seguintes condições:

- (i)  $p \lim_{T \rightarrow \infty} \widehat{\boldsymbol{\theta}}_T = \boldsymbol{\theta}_0$ ,
- (ii)  $\sqrt{T} \mathbf{g}(\boldsymbol{\theta}_0, \mathbb{Y}_T) \overset{\bullet}{\sim} N(\mathbf{0}, \mathbf{S})$ ,
- (ii) para qualquer sucessão  $(\boldsymbol{\theta}_T^*)$  que verifique  $p \lim_{T \rightarrow \infty} \boldsymbol{\theta}_T^* = \boldsymbol{\theta}_0$ , tem-se

$$p \lim_{T \rightarrow \infty} \left( \frac{\partial \mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T)}{\partial \boldsymbol{\theta}'} \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_T^*} \right) \left[ = p \lim_{T \rightarrow \infty} \left( \frac{\partial \mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T)}{\partial \boldsymbol{\theta}'} \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right) \right] = \mathbf{D}',$$

onde  $\mathbf{D}'$  é uma matriz  $(r \times a)$  com colunas linearmente independentes.

Então

$$\sqrt{T} (\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \overset{\bullet}{\sim} N(\mathbf{0}, \mathbf{V}),$$

onde  $\mathbf{V} = \{\mathbf{D}\mathbf{S}^{-1}\mathbf{D}'\}^{-1}$ .

□

Com base nesta proposição, pode considerar-se que  $\widehat{\boldsymbol{\theta}}_T \overset{\bullet}{\sim} N\left(\boldsymbol{\theta}_0, \frac{\hat{\mathbf{V}}}{T}\right)$ , onde

$$\hat{\mathbf{V}} = \{\hat{\mathbf{D}}\hat{\mathbf{S}}^{-1}\hat{\mathbf{D}}'\}^{-1}. \quad (3.6.11)$$

A matriz  $\hat{\mathbf{S}}$  pode ser calculada de acordo com (3.6.7), enquanto que  $\hat{\mathbf{D}}'_T$  se pode obter por

$$\hat{\mathbf{D}}'_T = \left\{ \frac{\partial \mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T)}{\partial \boldsymbol{\theta}'} \bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_T} \right\}. \quad (3.6.12)$$

Resumindo, para obter o estimador *GMM*, devem seguir-se os passos:

Passo 1 Identificar um conjunto de condições de ortogonalidade da forma

$$E[\mathbf{h}(\boldsymbol{\theta}_0, \mathbf{w}_t)] = \mathbf{0}.$$

Passo 2 Obter uma estimativa inicial  $\hat{\boldsymbol{\theta}}_0$  resolvendo o problema

$$\min_{\boldsymbol{\theta}} \{[\mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T)]' \mathbf{I}_T [\mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T)]\}. \text{ Esta estimativa será usada para obter a estimativa inicial de } \hat{\mathbf{S}}_T,$$

$$\hat{\mathbf{S}}_T^0 = \frac{1}{T} \sum_{t=1}^T [\mathbf{h}(\hat{\boldsymbol{\theta}}_0, \mathbf{w}_t)] [\mathbf{h}(\hat{\boldsymbol{\theta}}_0, \mathbf{w}_t)]'.$$

Passo 3 O estimador *GMM* para  $\boldsymbol{\theta}$  é o vector  $\widehat{\boldsymbol{\theta}}_T$ , definido pela solução do problema

$$\min_{\boldsymbol{\theta}} \mathbf{Q}(\boldsymbol{\theta}, \mathbb{Y}_T) = \min_{\boldsymbol{\theta}} [\mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T)]' \hat{\mathbf{S}}_T^{-1} [\mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T)],$$

onde

$$\mathbf{g}(\boldsymbol{\theta}, \mathbb{Y}_T) = \frac{1}{T} \sum_{t=1}^T \mathbf{h}(\boldsymbol{\theta}, \mathbf{w}_t),$$

e partindo de soluções iniciais  $\hat{\boldsymbol{\theta}}_0$  e  $\hat{\mathbf{S}}_T^0$ .

De seguida explicar-se-á o processo tradicional de adaptação da estimação por *GMM* a um *SEM*, tendo em conta as especificidades deste tipo de modelos, pelo facto de ser definido com base num sistema de equações simultâneas.

Uma vez que o *GMM* permite proceder à estimação dos parâmetros de um modelo em situações gerais da estrutura de covariâncias dos erros, passa a assumir-se a hipótese de que os erros de cada equação são heterocedásticos em relação ao tempo, ou seja, que se tem

$$E[e_{it}e_{it}] = \sigma_{it}^2 \text{ e } E[e_{it}e_{jt}] = \sigma_{ijt}.$$

A hipótese adoptada inicialmente, de que os erros aleatórios do modelo são gerados por um processo multivariado estacionário, deixa de ser válida, bastando admitir a existência de uma distribuição de equilíbrio que, no caso particular da homocedasticidade, coincide com a distribuição estacionária do processo anteriormente assumida. Observe-se que, neste caso, de acordo com (2.2.5) e (2.2.7), se tem que

$$\begin{aligned} E[\mathbf{e}_i \mathbf{e}_i'] &= \text{diag} \begin{bmatrix} \sigma_{i1}^2 & \sigma_{i2}^2 & \cdots & \sigma_{iT}^2 \end{bmatrix} = \boldsymbol{\Omega}_i, \\ \text{e } E[\mathbf{e}_i \mathbf{e}_j'] &= \text{diag} \begin{bmatrix} \sigma_{ij1} & \sigma_{ij2} & \cdots & \sigma_{ijT} \end{bmatrix} = \boldsymbol{\Omega}_{ij}. \end{aligned}$$

Efectuou-se algum trabalho no sentido de tentar exprimir a matriz  $\boldsymbol{\Omega}$  sob a forma de um produto de Kronecker, no entanto, como estes blocos são diagonais mas não são idênticos, não foi possível encontrar uma tal decomposição, pelo que a matriz  $\boldsymbol{\Omega}$  não se simplifica na forma referida em (3.2.3). Logo, o estudo prosseguiu com  $\boldsymbol{\Omega}$  na forma mais geral, definida em (3.2.2),

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Omega}_1 & \boldsymbol{\Omega}_{12} & \cdots & \boldsymbol{\Omega}_{1M} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_2 & \cdots & \boldsymbol{\Omega}_{2M} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\Omega}_{M1} & \boldsymbol{\Omega}_{M2} & \cdots & \boldsymbol{\Omega}_M \end{bmatrix}.$$

Na secção seguinte, apresenta-se o processo de estimação dos parâmetros por *GMM* separadamente para cada equação, método que será designado por *GMMIL*. Numa outra secção, estimam-se por *GMM* conjuntamente os parâmetros de todas as equações do modelo, método que será designado por *GMMIC*.

### 3.7 Estimadores *GMM* com informação limitada

Segundo o método da informação limitada, procede-se à estimação dos parâmetros de cada equação, separadamente. Note-se que, neste caso, não se consideram as correlações existentes entre os erros das diferentes equações. Consequente, os blocos que não pertencem à diagonal principal da matriz  $\Omega$  referida em (3.2.2) são nulos; isto é,

$$\Omega_{ij} = 0, \text{ se } i \neq j.$$

#### 3.7.1 Estimador *GMMIL*

Pretende-se estimar os parâmetros de cada equação  $i$ , a qual é da forma  $\mathbf{y}_i = \mathbf{Z}_i \boldsymbol{\delta}_i + \mathbf{e}_i$ , com  $\mathbf{Z}_i = \begin{bmatrix} \mathbf{Y}_i & \mathbf{X}_i \end{bmatrix}$ . Considere-se que  $\mathbf{X}$ , a matriz dos valores das variáveis exógenas, é correlacionada com  $\mathbf{Z}_i$ , uma vez que as colunas de  $\mathbf{X}_i$  também são colunas de  $\mathbf{X}$ . No entanto,  $\mathbf{X}$  não é correlacionada com  $\mathbf{e}_i$ , de acordo com HFE.6, ou seja, tem-se que  $E[\mathbf{X}'\mathbf{e}_i] = \mathbf{0}$ . Verificam-se, assim, as equações seguintes, designadas por condições de ortogonalidade,

$$E[\mathbf{X}'\mathbf{e}_i] = \mathbf{0} \Leftrightarrow E[\mathbf{X}'(\mathbf{y}_i - \mathbf{Z}_i \boldsymbol{\delta}_i)] = \mathbf{0}.$$

As condições amostrais correspondentes, também chamadas condições dos momentos amostrais, são obtidas quando se consideram os valores observados das variáveis, e são da forma

$$\frac{1}{T} \mathbf{X}' (\mathbf{y}_i - \mathbf{Z}_i \hat{\boldsymbol{\delta}}_i) = \mathbf{0} \Leftrightarrow \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t (y_{it} - \mathbf{z}_{it} \hat{\boldsymbol{\delta}}_i) = 0. \quad (3.7.1)$$

O estimador *GMM* para  $\boldsymbol{\delta}_i$  é obtido como solução desta equação matricial de momentos amostrais. Em relação ao procedimento geral de obtenção do estimador *GMM*, a equação (3.7.1) corresponde à estabelecida em (3.6.5); note-se ainda que, tendo presente os passos 1, 2 e 3 da secção anterior que resumem a obtenção das estimativas por

*GMM*, neste caso particular, se tem

$$\begin{aligned} \mathbf{w}_t &= (y_{it}, \mathbf{Z}'_{it}, \mathbf{X}'_t)', \boldsymbol{\theta} = \boldsymbol{\delta}_i, \\ \mathbf{h}(\boldsymbol{\delta}_i) &= \mathbf{x}_t (y_{it} - \mathbf{Z}_{it}\boldsymbol{\delta}_i), \\ \text{e } \mathbf{g}(\boldsymbol{\delta}_i) &= \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t (y_{it} - \mathbf{Z}_{it}\boldsymbol{\delta}_i). \end{aligned} \quad (3.7.2)$$

Se a equação  $i$  for exactamente identificada, a equação matricial (3.7.1) corresponde a um sistema em que o número de equações é igual ao número de incógnitas, sendo este último igual ao número componentes de  $\boldsymbol{\delta}_i$ , ( $k = k_i + M_i$ ). Deste modo, o sistema (3.7.1) tem uma solução única, dada por

$$\hat{\boldsymbol{\delta}}_{i,GMM} = [\mathbf{X}'\mathbf{Z}_i]^{-1} \mathbf{X}'\mathbf{y}_i.$$

Relembrando os estimadores de mínimos quadrados, conclui-se então que, se a equação  $i$  for exactamente identificada, o estimador *GMM* de informação limitada coincide com os estimadores *ILS* e *2SLS*, ou seja, neste caso, tem-se que  $\hat{\boldsymbol{\delta}}_{i,ILS} = \hat{\boldsymbol{\delta}}_{i,2SLS} = \hat{\boldsymbol{\delta}}_{i,GMM}$ .

Caso a equação  $i$  seja sobreidentificada, a equação (3.7.1) corresponde a um sistema em que o número de equações é superior ao número de parâmetros desconhecidos ( $k > k_i + M_i$ ). Assim, o método *GMM* já não estabelece que todas as condições de ortogonalidade amostrais sejam nulas. O estimador *GMM* de  $\boldsymbol{\delta}_i$  pode então ser obtido calculando a solução do problema

$$\min_{\boldsymbol{\delta}_i} \left\{ [T^{-1}\mathbf{X}'(\mathbf{y}_i - \mathbf{Z}_i\boldsymbol{\delta}_i)]' \mathbf{W}_i [T^{-1}\mathbf{X}'(\mathbf{y}_i - \mathbf{Z}_i\boldsymbol{\delta}_i)] \right\}$$

ou ainda, tendo em conta que o termo em  $T$  não vai ter efeito na solução do problema,

$$\min_{\boldsymbol{\delta}_i} \left\{ [\mathbf{X}'(\mathbf{y}_i - \mathbf{Z}_i\boldsymbol{\delta}_i)]' \mathbf{W}_i [\mathbf{X}'(\mathbf{y}_i - \mathbf{Z}_i\boldsymbol{\delta}_i)] \right\}.$$

Para resolver este problema de minimização, considera-se o sistema de estacionaridade,

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\delta}_i} \left[ (\mathbf{y}_i - \mathbf{Z}_i\hat{\boldsymbol{\delta}}_i)' \mathbf{X}\mathbf{W}_i\mathbf{X}' (\mathbf{y}_i - \mathbf{Z}_i\hat{\boldsymbol{\delta}}_i) \right] &= \mathbf{0} \\ \Leftrightarrow 2(-\mathbf{Z}'_i) \mathbf{X}\mathbf{W}_i\mathbf{X}' (\mathbf{y}_i - \mathbf{Z}_i\hat{\boldsymbol{\delta}}_i) &= \mathbf{0} \\ \Leftrightarrow \mathbf{Z}'_i\mathbf{X}\mathbf{W}_i\mathbf{X}' (\mathbf{y}_i - \mathbf{Z}_i\hat{\boldsymbol{\delta}}_i) &= \mathbf{0}, \end{aligned}$$

cuja solução é

$$\hat{\boldsymbol{\delta}}_{i,GMM} = [\mathbf{Z}'_i\mathbf{X}\mathbf{W}_i\mathbf{X}'\mathbf{Z}_i]^{-1} \mathbf{Z}'_i\mathbf{X}\mathbf{W}_i\mathbf{X}'\mathbf{y}_i. \quad (3.7.3)$$

De acordo com Hansen (1982), a matriz de pesos óptima, ou seja, aquela que permite obter o estimador *GMM* assintoticamente mais eficiente, é, neste caso,

$$\mathbf{W}_i = \left\{ V \left[ T^{-\frac{1}{2}} \mathbf{X}' \mathbf{e}_i \right] \right\}^{-1}.$$

Como

$$\begin{aligned} V \left[ T^{-\frac{1}{2}} \mathbf{X}' \mathbf{e}_i \right] &= E \left[ T^{-1} \mathbf{X}' \mathbf{e}_i \mathbf{e}_i' \mathbf{X} \right] \\ &= T^{-1} \mathbf{X}' E \left[ \mathbf{e}_i \mathbf{e}_i' \right] \mathbf{X} \\ &= T^{-1} \mathbf{X}' \boldsymbol{\Omega}_i \mathbf{X}, \end{aligned}$$

vem que

$$\mathbf{W}_i = \left[ T^{-1} \mathbf{X}' \boldsymbol{\Omega}_i \mathbf{X} \right]^{-1}.$$

Substituindo  $\mathbf{W}_i$  na expressão do estimador (3.7.3) e supondo  $\boldsymbol{\Omega}_i$  conhecida, obtém-se o estimador *GMM* para  $\boldsymbol{\delta}_i$

$$\hat{\boldsymbol{\delta}}_{i,GMM} = \left[ \mathbf{Z}_i' \mathbf{X} (\mathbf{X}' \boldsymbol{\Omega}_i \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z}_i \right]^{-1} \mathbf{Z}_i' \mathbf{X} (\mathbf{X}' \boldsymbol{\Omega}_i \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}_i. \quad (3.7.4)$$

De acordo com (3.6.11), a matriz de covariâncias assintótica do estimador é

$$VA \left[ \hat{\boldsymbol{\delta}}_{i,GMM} \right] = T^{-1} [\mathbf{D} \mathbf{W}_i \mathbf{D}']^{-1},$$

onde  $\mathbf{D}' = \left\{ \frac{\partial \mathbf{g}(\boldsymbol{\delta}_i)}{\partial \boldsymbol{\delta}_i'} \right\}$ . Atendendo a (3.7.2) e (3.6.12), pode-se obter

$$\begin{aligned} \mathbf{D}' &= \left\{ \frac{\partial \mathbf{g}(\boldsymbol{\delta}_i)}{\partial \boldsymbol{\delta}_i'} \right\} \Leftrightarrow \mathbf{D}' = \frac{\partial (\mathbf{X}' \mathbf{e}_i)}{\partial \boldsymbol{\delta}_i'} \\ &\Leftrightarrow \mathbf{D}' = \frac{\partial (\mathbf{X}' \mathbf{y}_i - \mathbf{X}' \mathbf{Z}_i \boldsymbol{\delta}_i)}{\partial \boldsymbol{\delta}_i'} \\ &\Leftrightarrow \mathbf{D}' = \mathbf{X}' \mathbf{Z}_i, \end{aligned}$$

donde vem que

$$VA \left[ \hat{\boldsymbol{\delta}}_{i,GMM} \right] = T^{-1} [\mathbf{Z}_i' \mathbf{X} \mathbf{W}_i \mathbf{X}' \mathbf{Z}_i]^{-1}.$$

No caso geral,  $\boldsymbol{\Omega}_i$  é desconhecida e é necessário estimar a matriz de pesos óptima. Tradicionalmente, recorre-se ao estimador proposto por White (1980), para a matriz de covariâncias no caso heterocedástico, ou seja,

$$\hat{\mathbf{W}}_i = \left[ T^{-1} \mathbf{X}' \hat{\boldsymbol{\Omega}}_i \mathbf{X} \right]^{-1},$$



onde

$$\hat{\Omega}_i = \text{diag} \begin{bmatrix} \hat{e}_{1i}^2 & \hat{e}_{2i}^2 & \cdots & \hat{e}_{Ti}^2 \end{bmatrix} \text{ e } \hat{e}_{ti} = y_{ti} - \mathbf{Z}_{ti}' \hat{\boldsymbol{\delta}}_{i,2SLS}.$$

Sabe-se que o estimador de White é consistente para estimar  $\mathbf{W}_i$ , mas chama-se a atenção para o facto de cada elemento  $e_{ti}^2$  de  $\Omega_i$  ser estimado apenas com base num único resíduo,  $\hat{e}_{ti}$ . Substituindo  $\Omega_i$  por  $\hat{\Omega}_i$  em (3.7.4), obtém-se a forma do estimador *GMM* para , no caso de  $\Omega_i$  ser desconhecida:

$$\hat{\boldsymbol{\delta}}_{i,GMM}(\hat{\Omega}_i) = \left[ \mathbf{Z}_i' \mathbf{X} (\mathbf{X}' \hat{\Omega}_i \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z}_i \right]^{-1} \mathbf{Z}_i' \mathbf{X} (\mathbf{X}' \hat{\Omega}_i \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}_i. \quad (3.7.5)$$

Admitindo as habituais condições de regularidade, nomeadamente, que

$$p \lim_{T \rightarrow \infty} [T^{-1} \mathbf{X}' \mathbf{e}_i] = \mathbf{0},$$

e que as matrizes

$$\begin{aligned} p \lim_{T \rightarrow \infty} [T^{-1} (\mathbf{X}' \Omega_i \mathbf{X})] &= \Sigma_{\mathbf{X} \Omega_i \mathbf{X}} \\ p \lim_{T \rightarrow \infty} [T^{-1} \mathbf{X}' \mathbf{Z}_i] &= \Sigma_{\mathbf{X} \mathbf{Z}_i} \end{aligned}$$

são finitas e não singulares, mostra-se que, quando  $\Omega_i$  é estimada por  $\hat{\Omega}_i$ , o estimador  $\hat{\boldsymbol{\delta}}_{i,GMM}(\hat{\Omega}_i)$ , definido por (3.7.5), goza das seguintes propriedades (as demonstrações podem ser vistas em Mittelhammer *et al.*, 2000):

PGMMIL.1  $\hat{\boldsymbol{\delta}}_{i,GMM}$  é um estimador consistente para  $\boldsymbol{\delta}_i$ , isto é,

$$p \lim_{T \rightarrow \infty} \hat{\boldsymbol{\delta}}_{i,GMM}(\hat{\Omega}_i) = \boldsymbol{\delta}_i.$$

PGMMIL.2  $\hat{\boldsymbol{\delta}}_{i,GMM}$  tem distribuição assintótica normal definida por:

$$\hat{\boldsymbol{\delta}}_{i,GMM}(\hat{\Omega}_i) \stackrel{\bullet}{\sim} N(\boldsymbol{\delta}_i, \Sigma_{\mathbf{Z}_i \mathbf{X}} \Sigma_{\mathbf{X} \Omega_i \mathbf{X}}^{-1} \Sigma_{\mathbf{X} \mathbf{Z}_i}).$$

PGMMIL.3 Um estimador consistente da variância assintótica de  $\hat{\boldsymbol{\delta}}_{i,GMM}$  é o seguinte

$$\widehat{VA} \left[ \hat{\boldsymbol{\delta}}_{i,GMM}(\hat{\Omega}_i) \right] = T^{-1} \left[ \mathbf{Z}_i' \mathbf{X} (T^{-1} \mathbf{X}' \hat{\Omega}_i \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z}_i \right]^{-1}.$$

Importa ainda obter a forma do estimador *GMMIL* para todas as equações do sistema, uma vez que até agora se vem a considerar apenas a estimação da equação  $i$ .

Para obter o estimador *GMMIL* para o vector  $\boldsymbol{\delta}$  no seu todo, substitui-se cada bloco  $\boldsymbol{\delta}_i$  pelo respectivo estimador  $\hat{\boldsymbol{\delta}}_{i,GMM}$  na forma (3.7.4), resultando em

$$\begin{aligned}\hat{\boldsymbol{\delta}}_{GMMIL} &= \begin{bmatrix} \hat{\boldsymbol{\delta}}'_{1,GMM} & \dots & \hat{\boldsymbol{\delta}}'_{M,GMM} \end{bmatrix}' \\ &= \begin{bmatrix} [\mathbf{Z}'_1 \mathbf{X} (\mathbf{X}' \boldsymbol{\Omega}_1 \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z}_1]^{-1} \mathbf{Z}'_1 \mathbf{X} (\mathbf{X}' \boldsymbol{\Omega}_1 \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}_1 \\ \vdots \\ [\mathbf{Z}'_M \mathbf{X} (\mathbf{X}' \boldsymbol{\Omega}_M \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z}_M]^{-1} \mathbf{Z}'_M \mathbf{X} (\mathbf{X}' \boldsymbol{\Omega}_M \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}_M \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{Z}'_1 \mathbf{X} (\mathbf{X}' \boldsymbol{\Omega}_1 \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z}_1 & & \\ & \ddots & \\ & & \mathbf{Z}'_M \mathbf{X} (\mathbf{X}' \boldsymbol{\Omega}_M \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z}_M \end{bmatrix}^{-1} \\ &\times \begin{bmatrix} \mathbf{Z}'_1 \mathbf{X} (\mathbf{X}' \boldsymbol{\Omega}_1 \mathbf{X})^{-1} & & \\ & \ddots & \\ & & \mathbf{Z}'_M \mathbf{X} (\mathbf{X}' \boldsymbol{\Omega}_M \mathbf{X})^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{X}' \mathbf{y}_1 \\ \vdots \\ \mathbf{X}' \mathbf{y}_M \end{bmatrix}.\end{aligned}$$

Note-se que os blocos da matriz  $\boldsymbol{\Omega}$  que não pertencem à diagonal principal são nulos, e a matriz  $\boldsymbol{\Omega}$  simplifica-se na forma

$$\boldsymbol{\Omega}_{IL} = \begin{bmatrix} \boldsymbol{\Omega}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Omega}_M \end{bmatrix}.$$

Usando o produto de Kronecker e  $\boldsymbol{\Omega}_{IL}$ , a expressão de  $\hat{\boldsymbol{\delta}}_{GMMIL}$  simplifica-se em

$$\begin{aligned}\hat{\boldsymbol{\delta}}_{GMMIL} &= \left\{ \mathbf{Z}' (\mathbf{I}_M \otimes \mathbf{X}) [(\mathbf{I}_M \otimes \mathbf{X}') \boldsymbol{\Omega}_{IL} (\mathbf{I}_M \otimes \mathbf{X})]^{-1} (\mathbf{I}_M \otimes \mathbf{X}') \mathbf{Z} \right\}^{-1} \\ &\times \mathbf{Z}' (\mathbf{I}_M \otimes \mathbf{X}) [(\mathbf{I}_M \otimes \mathbf{X}') \boldsymbol{\Omega}_{IL} (\mathbf{I}_M \otimes \mathbf{X})]^{-1} (\mathbf{I}_M \otimes \mathbf{X}') \mathbf{y}.\end{aligned}$$

A expressão anterior pode ainda ser mais compactada, definindo

$$\mathbf{C} = (\mathbf{I}_M \otimes \mathbf{X}') \mathbf{Z}, \boldsymbol{\Omega}_{IL}^* = (\mathbf{I}_M \otimes \mathbf{X}') \boldsymbol{\Omega}_{IL} (\mathbf{I}_M \otimes \mathbf{X}) \text{ e } \mathbf{y}^* = (\mathbf{I}_M \otimes \mathbf{X}') \mathbf{y}. \quad (3.7.6)$$

Deste modo, vem que

$$\hat{\boldsymbol{\delta}}_{GMMIL} = [\mathbf{C}' (\boldsymbol{\Omega}_{IL}^*)^{-1} \mathbf{C}]^{-1} \mathbf{C}' (\boldsymbol{\Omega}_{IL}^*)^{-1} \mathbf{y}^*. \quad (3.7.7)$$

### 3.8 Estimadores *GMM* com informação completa

Recorde-se que, com esta metodologia, se procede à estimação simultânea dos parâmetros de todas as equações do modelo. Neste caso, e ao admitir a hipótese de heterocedasticidade, vem para  $\mathbf{\Omega} = \mathbf{V}[\mathbf{e}] = E[\mathbf{e}\mathbf{e}']$  a forma mais geral (3.2.2), ou seja,

$$\mathbf{\Omega} = \begin{bmatrix} \mathbf{\Omega}_1 & \cdots & \mathbf{\Omega}_{1M} \\ \vdots & \ddots & \vdots \\ \mathbf{\Omega}_{M1} & \cdots & \mathbf{\Omega}_M \end{bmatrix},$$

onde cada  $\mathbf{\Omega}_{ij}$  é um bloco diagonal com elementos diferentes, do tipo

$\text{diag} \begin{bmatrix} \sigma_{ij1} & \sigma_{ij2} & \cdots & \sigma_{ijT} \end{bmatrix}$ , para  $i, j = 1, \dots, M$ , sem que à partida se possam admitir simplificações desta matriz.

#### 3.8.1 Estimador *GMMIC*

Para estimar simultaneamente os parâmetros de todas as equações, retome-se o modelo completo, na forma (3.4.6), isto é

$$(\mathbf{I}_M \otimes \mathbf{X}') \mathbf{y} = (\mathbf{I}_M \otimes \mathbf{X}') \mathbf{Z} \boldsymbol{\delta} + (\mathbf{I}_M \otimes \mathbf{X}') \mathbf{e}.$$

Em consequência da hipótese HFE.6 admitida para o modelo, a qual estabelece que as variáveis exógenas e os erros aleatórios do modelo são não correlacionados, as condições de ortogonalidade são agora

$$E[(\mathbf{I}_M \otimes \mathbf{X}') \mathbf{e}] = \mathbf{0}.$$

Para obter o estimador *GMM* para  $\boldsymbol{\delta}$ , é necessário obter a correspondente equação dos momentos amostrais

$$\frac{1}{T} (\mathbf{I}_M \otimes \mathbf{X}') (\mathbf{y} - \mathbf{Z} \boldsymbol{\delta}) = \mathbf{0},$$

expressa em termos dos valores observados das variáveis, e que corresponde a um sistema formado por  $MK$  equações.

O estimador *GMMIC* é definido pelo vector de parâmetros  $\boldsymbol{\delta}$  que faz com que as condições dos momentos amostrais estejam tão próximas de zero quanto possível. Assim, no caso geral,  $\hat{\boldsymbol{\delta}}_{GMMIC}$  corresponde à solução do problema

$$\min_{\boldsymbol{\delta}} \left\{ [T^{-1} (\mathbf{I}_M \otimes \mathbf{X}') \mathbf{e}]' \mathbf{W} [T^{-1} (\mathbf{I}_M \otimes \mathbf{X}') \mathbf{e}] \right\}, \quad (3.8.1)$$

ou  $\min_{\delta} \{[(\mathbf{I}_M \otimes \mathbf{X}') \mathbf{e}]' \mathbf{W} [(\mathbf{I}_M \otimes \mathbf{X}') \mathbf{e}]\}$ , supondo  $\mathbf{W}$  conhecido. Para encontrar a solução deste problema, resolve-se o sistema de estacionaridade,

$$\begin{aligned} \frac{\partial}{\partial \delta} [(\mathbf{y} - \mathbf{Z}\delta)' (\mathbf{I}_M \otimes \mathbf{X}) \mathbf{W} (\mathbf{I}_M \otimes \mathbf{X}') (\mathbf{y} - \mathbf{Z}\delta)] &= \mathbf{0} \\ \Leftrightarrow 2(-\mathbf{Z}') (\mathbf{I}_M \otimes \mathbf{X}) \mathbf{W} (\mathbf{I}_M \otimes \mathbf{X}') (\mathbf{y} - \mathbf{Z}\delta) &= \mathbf{0} \\ \Leftrightarrow \mathbf{Z}' (\mathbf{I}_M \otimes \mathbf{X}) \mathbf{W} (\mathbf{I}_M \otimes \mathbf{X}') (\mathbf{y} - \mathbf{Z}\delta) &= \mathbf{0}, \\ \Leftrightarrow \mathbf{Z}' (\mathbf{I}_M \otimes \mathbf{X}) \mathbf{W} (\mathbf{I}_M \otimes \mathbf{X}') \mathbf{y} - \mathbf{Z}' (\mathbf{I}_M \otimes \mathbf{X}) \mathbf{W} (\mathbf{I}_M \otimes \mathbf{X}') \mathbf{Z}\delta &= \mathbf{0}. \end{aligned}$$

Logo, supondo  $\mathbf{W}$  conhecido, a solução de (3.8.1) define o estimador

$$\hat{\delta}_{GMMIC} = [\mathbf{Z}' (\mathbf{I}_M \otimes \mathbf{X}) \mathbf{W} (\mathbf{I}_M \otimes \mathbf{X}') \mathbf{Z}]^{-1} \mathbf{Z}' (\mathbf{I}_M \otimes \mathbf{X}) \mathbf{W} (\mathbf{I}_M \otimes \mathbf{X}') \mathbf{y}. \quad (3.8.2)$$

No sentido de obter o estimador *GMM* mais eficiente, Hansen (1982) mostrou que a matriz de pesos óptima deve ser

$$\mathbf{W} = \{V [T^{-1/2} (\mathbf{I}_M \otimes \mathbf{X}') \mathbf{e}]\}^{-1}.$$

Como neste caso,

$$\begin{aligned} V [T^{-1/2} (\mathbf{I}_M \otimes \mathbf{X}') \mathbf{e}] &= T^{-1} E [(\mathbf{I}_M \otimes \mathbf{X}') \mathbf{e} \mathbf{e}' (\mathbf{I}_M \otimes \mathbf{X})] \\ &= T^{-1} (\mathbf{I}_M \otimes \mathbf{X}') \mathbf{\Omega} (\mathbf{I}_M \otimes \mathbf{X}), \end{aligned}$$

vem que

$$\mathbf{W} = [T^{-1} (\mathbf{I}_M \otimes \mathbf{X}') \mathbf{\Omega} (\mathbf{I}_M \otimes \mathbf{X})]^{-1}.$$

Suponha-se de momento que  $\mathbf{\Omega}$  é conhecida. Substituindo esta matriz de pesos na expressão (3.8.2), viria

$$\begin{aligned} \hat{\delta}_{GMMIC} &= \left[ \mathbf{Z}' (\mathbf{I}_M \otimes \mathbf{X}) [(\mathbf{I}_M \otimes \mathbf{X}') \mathbf{\Omega} (\mathbf{I}_M \otimes \mathbf{X})]^{-1} (\mathbf{I}_M \otimes \mathbf{X}') \mathbf{Z} \right]^{-1} \\ &\times \mathbf{Z}' (\mathbf{I}_M \otimes \mathbf{X}) [(\mathbf{I}_M \otimes \mathbf{X}') \mathbf{\Omega} (\mathbf{I}_M \otimes \mathbf{X})]^{-1} (\mathbf{I}_M \otimes \mathbf{X}') \mathbf{y}. \end{aligned} \quad (3.8.3)$$

Usando  $\mathbf{C}$  e  $\mathbf{y}^*$  definidos em (3.7.6), e considerando  $\mathbf{\Omega}^* = (\mathbf{I}_M \otimes \mathbf{X}') \mathbf{\Omega} (\mathbf{I}_M \otimes \mathbf{X})$ , torna-se possível apresentar o estimador numa forma mais compacta,

$$\hat{\delta}_{GMMIC} = [\mathbf{C}' (\mathbf{\Omega}^*)^{-1} \mathbf{C}]^{-1} \mathbf{C}' (\mathbf{\Omega}^*)^{-1} \mathbf{y}^*. \quad (3.8.4)$$

Sendo a matriz  $\Omega$  desconhecida, então  $\Omega^*$  também o é, sendo necessário estimá-la previamente. O processo usual é recorrer ao estimador consistente da matriz de covariâncias para o caso heterocedástico proposto em White (1980), com base nos resíduos obtidos por *2SLS*, isto é,

$$\hat{\mathbf{W}} = \left[ T^{-1} (\mathbf{I}_M \otimes \mathbf{X}') \hat{\Omega} (\mathbf{I}_M \otimes \mathbf{X}) \right]^{-1},$$

onde

$$\begin{aligned} \hat{\Omega} &= \left[ \hat{\Omega}_{ij} \right]_{i,j=1,\dots,M}, \\ \hat{\Omega}_i &= \text{diag} \left[ \hat{e}_{1i}^2 \quad \hat{e}_{2i}^2 \quad \cdots \quad \hat{e}_{Ti}^2 \right], \\ \hat{\Omega}_{ij} &= \text{diag} \left[ \hat{e}_{1i}\hat{e}_{1j} \quad \hat{e}_{2i}\hat{e}_{2j} \quad \cdots \quad \hat{e}_{Ti}\hat{e}_{Tj} \right], \\ \text{e } \hat{e}_{ti} &= y_{ti} - \mathbf{Z}_{ti}\hat{\boldsymbol{\delta}}_{i,2SLS}. \end{aligned}$$

Uma vez mais realça-se o facto do estimador de White usar um único resíduo para estimar cada elemento de  $\Omega$ . Para apresentar as propriedades do estimador, admitam-se as seguintes condições

$$(i) \quad p \lim [T^{-1} (\mathbf{I}_M \otimes \mathbf{X}') \mathbf{e}] = 0,$$

e ainda que as seguintes matrizes são finitas e não singulares,

$$(ii) \quad p \lim [T^{-1} ((\mathbf{I}_M \otimes \mathbf{X}') \Omega (\mathbf{I} \otimes \mathbf{X}))] = \Sigma_{\mathbf{X}\Omega\mathbf{X}},$$

$$(iii) \quad p \lim [T^{-1} (\mathbf{I}_M \otimes \mathbf{X}') \mathbf{Z}] = \Sigma_{\mathbf{XZ}}.$$

Prova-se em Mittelhammer *et al.* (2000) que o estimador que se obtém substituindo  $\Omega$  por  $\hat{\Omega}$  em (3.8.3), e que será denotado por  $\hat{\boldsymbol{\delta}}_{GMMIC}(\hat{\Omega})$ , verifica as seguintes propriedades:

PGMMIC.1  $\hat{\boldsymbol{\delta}}_{GMMIC}$  é um estimador consistente para  $\boldsymbol{\delta}$ , isto é,

$$p \lim \hat{\boldsymbol{\delta}}_{GMMIC}(\hat{\Omega}) = \boldsymbol{\delta}.$$

PGMMIC.2  $\hat{\boldsymbol{\delta}}_{GMMIC}$  tem distribuição assintótica normal definida por:

$$\hat{\boldsymbol{\delta}}_{GMMIC}(\hat{\Omega}) \overset{\circ}{\sim} N(\boldsymbol{\delta}, \Sigma_{\mathbf{ZX}} \Sigma_{\mathbf{X}\Omega\mathbf{X}}^{-1} \Sigma_{\mathbf{XZ}}).$$

PGMMIC.3 O estimador da matriz de covariâncias assintótica de  $\hat{\delta}_{GMMIC}(\hat{\Omega})$  pode ser expresso na forma

$$\widehat{VA} \left[ \hat{\delta}_{GMMIC}(\hat{\Omega}) \right] = T^{-1} \left[ \mathbf{Z}' (\mathbf{I}_M \otimes \mathbf{X}) \left[ T^{-1} (\mathbf{I}_M \otimes \mathbf{X}') \hat{\Omega} (\mathbf{I}_M \otimes \mathbf{X}) \right]^{-1} (\mathbf{I}_M \otimes \mathbf{X}') \mathbf{Z} \right]^{-1}.$$

### 3.9 Comparação das matrizes de covariâncias assintóticas dos estimadores *GMMIL* e *GMMIC*

Na secção dedicada ao estudo dos estimadores de mínimos quadrados, referiu-se que o método de informação limitada e o método de informação completa tinham sido comparados, tendo-se concluído que teoricamente o método de informação completa é assintoticamente mais eficiente. Para o caso do método dos momentos generalizado, esse estudo era inexistente. A presente secção contém essa comparação desenvolvida originalmente durante o presente trabalho. Em linhas gerais, a comparação dos estimadores baseou-se no estudo da eficiência relativa assintótica entre os estimadores  $\hat{\delta}_{GMMIL}$  e  $\hat{\delta}_{GMMIC}$ . A conclusão foi idêntica à referida na **Secção 3.5**, isto é, que, teoricamente, o estimador *GMMIC* é assintoticamente mais eficiente do que *GMMIL*.

O resultado alcançado foi possível depois de se terem encontrado novas representações das matrizes de covariâncias dos estimadores  $\hat{\delta}_{GMMIL}$  e  $\hat{\delta}_{GMMIC}$ . Estas representações serão apresentadas nas **Proposições 3.9.1** e **3.9.3**, não se encontram na literatura e facilitaram a comparação das matrizes, como se verá no **Teorema 3.9.5**.

Passa a apresentar-se a obtenção da nova representação da variância de  $\hat{\delta}_{GMMIL}$ . O resultado expresso na **Proposição 3.9.1** foi obtido pela autora, não se encontrando publicado na literatura. No entanto, em Mittelhammer *et al.* (2000) é apresentado um resultado correspondente para o estimador dos parâmetros de cada equação, embora para o caso particular do modelo homocedástico e sem divulgação do processo de cálculo.

**Proposição 3.9.1.** A matriz de covariâncias do estimador *GMMIL* definido em (3.7.7) é da forma

$$V \left[ \hat{\delta}_{GMMIL} \right] = \left[ \mathbf{C}' (\Omega_{IL}^*)^{-1} \mathbf{C} \right]^{-1}, \quad (3.9.1)$$

onde  $C$  e  $\Omega_{IL}^*$  estão definidas em 3.7.6 e supondo que  $\Omega_{IL}^*$  é conhecida.

*Demonstração.* A matriz de covariâncias do estimador  $GMMIL$  na forma (3.7.7) será

$$\begin{aligned} V \left[ \hat{\delta}_{GMMIL} \right] &= \left\{ [C' (\Omega_{IL}^*)^{-1} C]^{-1} C' (\Omega_{IL}^*)^{-1} \right\} V [y^*] \left\{ [C' (\Omega_{IL}^*)^{-1} C]^{-1} C' (\Omega_{IL}^*)^{-1} \right\}' \\ &= \left\{ [C' (\Omega_{IL}^*)^{-1} C]^{-1} C' (\Omega_{IL}^*)^{-1} \right\} V [(I \otimes X') e] \\ &\quad \times \left\{ [C' (\Omega_{IL}^*)^{-1} C]^{-1} C' (\Omega_{IL}^*)^{-1} \right\}' \\ &= \left\{ [C' (\Omega_{IL}^*)^{-1} C]^{-1} C' (\Omega_{IL}^*)^{-1} \right\} [(I \otimes X') \Omega_{IL} (I \otimes X)] \\ &\quad \times \left\{ [C' (\Omega_{IL}^*)^{-1} C]^{-1} C' (\Omega_{IL}^*)^{-1} \right\}'. \end{aligned}$$

Efectuando alguns cálculos matriciais, esta expressão simplifica-se em

$$\begin{aligned} V \left[ \hat{\delta}_{GMMIL} \right] &= [C' (\Omega_{IL}^*)^{-1} C]^{-1} \\ &= \left\{ Z' (I \otimes X) [(I \otimes X') \Omega_{IL} (I \otimes X)]^{-1} (I \otimes X') Z \right\}^{-1}. \end{aligned}$$

□

**Corolário 3.9.2.** Supondo que se verifica a condição  $p \lim [T^{-1} ((I_M \otimes X') \Omega_{IL} (I_M \otimes X))] = \Sigma_{X \Omega_{IL} X}$ , então vem que

$$VA \left[ \hat{\delta}_{GMMIL} \right] = p \lim V \left[ \hat{\delta}_{GMMIL} \right] = [T^{-1} \Sigma_{ZX} \Sigma_{X \Omega_{IL} X}^{-1} \Sigma_{XZ}]^{-1}.$$

□

Na **Proposição 3.9.3** apresenta-se o desenvolvimento de uma representação compacta da matriz de covariâncias do estimador  $GMMIC$ . Note-se que em Mittelhammer *et al.* (2000), os autores apresentam um resultado análogo, mas para o caso homocedástico e sem demonstração. No entanto, a representação que se propõe neste trabalho, para além de ser mais geral por contemplar o caso heterocedástico, torna-se mais adequada para comparar a eficiência dos estimadores.

**Proposição 3.9.3.** A matriz de covariâncias do estimador  $GMMIC$  definido em (3.8.4) é da forma

$$V \left[ \hat{\delta}_{GMMIC} \right] = [C' (\Omega^*)^{-1} C]^{-1}, \quad (3.9.2)$$

onde  $C$  é definida em (3.7.6), e  $\Omega^* = (I_M \otimes X') \Omega (I_M \otimes X)$  e admitindo que  $\Omega$  é conhecida.

*Demonstração.* Para determinar a matriz de covariâncias do estimador  $GMMIC$ , considerou-se o estimador na forma (3.8.4) e obteve-se

$$\begin{aligned} V \left[ \hat{\delta}_{GMMIC} \right] &= \left\{ \left[ \mathbf{C}' (\boldsymbol{\Omega}^*)^{-1} \mathbf{C} \right]^{-1} \mathbf{C}' (\boldsymbol{\Omega}^*)^{-1} \right\} V [\mathbf{y}^*] \left\{ \left[ \mathbf{C}' (\boldsymbol{\Omega}^*)^{-1} \mathbf{C} \right]^{-1} \mathbf{C}' (\boldsymbol{\Omega}^*)^{-1} \right\}' \\ &= \left\{ \left[ \mathbf{C}' (\boldsymbol{\Omega}^*)^{-1} \mathbf{C} \right]^{-1} \mathbf{C}' (\boldsymbol{\Omega}^*)^{-1} \right\} V [(\mathbf{I} \otimes \mathbf{X}') \mathbf{e}] \\ &\quad \times \left\{ \left[ \mathbf{C}' (\boldsymbol{\Omega}^*)^{-1} \mathbf{C} \right]^{-1} \mathbf{C}' (\boldsymbol{\Omega}^*)^{-1} \right\}' \\ &= \left\{ \left[ \mathbf{C}' (\boldsymbol{\Omega}^*)^{-1} \mathbf{C} \right]^{-1} \mathbf{C}' (\boldsymbol{\Omega}^*)^{-1} \right\} [(\mathbf{I} \otimes \mathbf{X}') \boldsymbol{\Omega} (\mathbf{I} \otimes \mathbf{X})] \\ &\quad \times \left\{ \left[ \mathbf{C}' (\boldsymbol{\Omega}^*)^{-1} \mathbf{C} \right]^{-1} \mathbf{C}' (\boldsymbol{\Omega}^*)^{-1} \right\}'. \end{aligned}$$

Após alguns cálculos matriciais, a expressão simplifica-se em

$$\begin{aligned} V \left[ \hat{\delta}_{GMMIC} \right] &= \left[ \mathbf{C}' (\boldsymbol{\Omega}^*)^{-1} \mathbf{C} \right]^{-1} \\ &= \left\{ \mathbf{Z}' (\mathbf{I} \otimes \mathbf{X}) [(\mathbf{I} \otimes \mathbf{X}') \boldsymbol{\Omega} (\mathbf{I} \otimes \mathbf{X})]^{-1} (\mathbf{I} \otimes \mathbf{X}') \mathbf{Z} \right\}^{-1}. \end{aligned}$$

□

**Corolário 3.9.4.** Admitindo as condições (ii) e (iii), tem-se que

$$VA \left[ \hat{\delta}_{GMMIC} \right] = p \lim V \left[ \hat{\delta}_{GMMIC} \right] = \left[ T^{-1} \boldsymbol{\Sigma}_{\mathbf{Z}\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X}\boldsymbol{\Omega}\mathbf{X}}^{-1} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{Z}} \right]^{-1}.$$

□

Na prática as matrizes  $\boldsymbol{\Omega}_{IL}$  e  $\boldsymbol{\Omega}$ , que intervêm na expressões das variâncias assintóticas dos estimadores, são desconhecidas e é necessário estimá-las. De acordo com o método tradicional, recorre-se ao estimador de White (White, 1980), conforme foi explicado anteriormente, para obter estimadores consistentes  $\hat{\boldsymbol{\Omega}}_{IL}$  e  $\hat{\boldsymbol{\Omega}}$ , respectivamente, os quais são inseridos nas expressões que definem  $\hat{\delta}_{GMMIL}$  e  $\hat{\delta}_{GMMIC}$  (usando, respectivamente, as expressões (3.7.7) e (3.8.4)).

Recordando o que se referiu para estimadores de mínimos quadrados, e para comparar os estimadores  $3SLS$  e  $2SLS$ , encontra-se em Judge *et al.* (1988) uma demonstração de que  $VA \left[ \hat{\delta}_{2SLS} \right] - VA \left[ \hat{\delta}_{3SLS} \right]$  é uma matriz semidefinida positiva. Logo, de acordo com o Critério.1, tem-se

$$VA \left[ \hat{\delta}_{2SLS} \right] \geq VA \left[ \hat{\delta}_{3SLS} \right],$$



pelo que se concluiu que  $\hat{\delta}_{3SLS}$  é assintoticamente mais eficiente do que  $\hat{\delta}_{2SLS}$ .

Para estabelecer uma comparação semelhante entre os estimadores  $GMMIC$  e  $GMMIL$ , estudou-se a eficiência relativa assintótica, usando diferentes critérios. Em primeiro lugar, procurou-se usar o Critério.1, pois é o critério mais forte para a comparação de matrizes. Infelizmente não foi possível confirmar que  $\mathbf{D} = VA \left[ \hat{\delta}_{GMMIL} \right] - VA \left[ \hat{\delta}_{GMMIC} \right]$  é semidefinida positiva.

Na impossibilidade de estabelecer resultados por esse critério, utilizou-se de seguida o Critério.2 (critério da variância generalizada), menos forte do que o anterior. De acordo com este último critério, a comparação de matrizes é efectuada a partir dos respectivos determinantes. Para isso foi necessário calcular os determinantes das matrizes  $VA \left[ \hat{\delta}_{GMMIL} \right]$  e  $VA \left[ \hat{\delta}_{GMMIC} \right]$ .

Usando as **Proposições 3.9.1** e **3.9.3**, foi possível concluir qual dos métodos é assintoticamente mais eficiente. Este resultado encontra-se no teorema que se segue. A demonstração é muito longa e foi, por isso, remetida para anexo, mas baseia-se nas seguintes linhas gerais: Trata-se de uma demonstração feita por indução matemática sobre o número  $M$  de equações do modelo. Para estabelecer a relação entre os determinantes, partiu-se de um sistema com duas equações (ou seja, com  $M = 2$ ). Para  $M = 2$  foi possível estabelecer uma desigualdade entre as matrizes de covariância, recorrendo a uma decomposição envolvendo matrizes triangulares por blocos, apresentada em Seber (1984). A partir dessa decomposição, o determinante de cada uma das matrizes foi expresso através de produtos de determinantes de matrizes auxiliares. De seguida, observou-se que as matrizes auxiliares eram matrizes de covariâncias ou matrizes de covariâncias parciais, de acordo com Mardia *et al.* (1979). Pelas propriedades básicas do coeficiente de correlação foi possível estabelecer a hipótese de indução. A generalização para um modelo com  $M$  equações, para qualquer  $M > 2$ , envolveu o mesmo tipo de técnicas.

**Teorema 3.9.5.** Considere-se um SEM na forma  $\mathbf{y} = \mathbf{Z}\delta + \mathbf{e}$ , definida em (3.4.2) e sejam  $\hat{\delta}_{GMMIL}$  e  $\hat{\delta}_{GMMIC}$  os estimadores de  $\delta$ , obtidos pelo Método dos Momentos Genaralizado com informação limitada e com informação completa, definidos, respectivamente, pelas expressões (3.7.7) e (3.8.4). De acordo com o Critério da variância generalizada (Critério.2), o estimador  $\hat{\delta}_{GMMIC}$  é assintoticamente mais eficiente do que

o estimador  $\hat{\delta}_{GMMIL}$ .

*Demonstração.* A demonstração pode ser consultada no Apêndice A.2.3.  $\square$

Este resultado evidencia a vantagem teórica do estimador  $\hat{\delta}_{GMMIC}$  sobre o estimador  $\hat{\delta}_{GMMIL}$ , traduzindo que o estimador de informação completa conduz a uma maior eficiência assintótica do que o estimador de informação limitada. Assim, em amostras de dimensão elevada, há vantagem em usar informação completa. No entanto, e à semelhança do que acontece com os estimadores de mínimos quadrados, em amostras de pequena dimensão essa vantagem não é significativa. Voltar-se-á a esta questão na secção seguinte.

### 3.10 Estudo empírico da eficiência relativa entre estimadores de informação limitada e de informação completa

Para ilustrar a aplicação dos métodos de estimação apresentados ao longo deste capítulo e avaliar a superioridade dos métodos de informação completa sobre os de informação limitada em estudos práticos e, portanto, em dimensão finita, utilizaram-se dois exemplos de modelos econométricos. Note-se que a modelação do relacionamento entre as variáveis provém da Teoria Económica e que o objectivo do presente estudo estatístico dos *SEM* diz respeito à estimação dos parâmetros e não à modelação econométrica.

De acordo com as **Secções 3.5 e 3.9**, os estimadores que são obtidos considerando simultaneamente todos os parâmetros do modelo (por *MIC*) são assintoticamente mais eficientes do que os estimadores que resultam de proceder à estimação dos parâmetros, separadamente para cada equação (por *MIL*). No entanto esta conclusão, que é válida em termos assintóticos, não fica patente de forma significativa em estudos de dimensão finita. Este comentário traduz as dúvidas levantadas por alguns autores (por exemplo, Greene, 2003 e Baltagi, 2008), que por diversas vezes têm questionado a superioridade dos métodos de informação completa. Esses autores advogam que os *MIL* dão melhores resultados em aplicações a casos reais, sem no entanto conseguirem explicar teoricamente a justificação para tal facto. Os estudos de simulação que se apresentam num

dos exemplos seguintes, ajudam a explicar os motivos que poderão estar na base das dúvidas referidas em relação à superioridade dos *MIC*. Nos dois exemplos apresentados nesta secção, verificou-se que as diferenças entre os determinantes das estimativas das matrizes de covariâncias dos estimadores de informação limitada e completa são sempre muito próximas de zero. Isto significa que, apesar dos *MIC* serem assintoticamente superiores, em dimensão finita os dois tipos de métodos são praticamente equivalentes. Assim, não é de estranhar que os *MIL* sejam frequentemente preferidos na prática, quer pela estrutura particular de cada modelo, quer pela maior simplicidade de manuseamento computacional dos *MIL*.

O estudo efectuado indicou que, para os casos considerados, em dimensões finitas e independentemente do método de estimação usado (método dos mínimos quadrados ou método dos momentos generalizado), não há diferenças significativas em termos da eficiência ao usar o estimador *MIL* ou o estimador *MIC*. A elevada dimensão das matrizes a manipular nos *MIC* não conduz a ganhos de eficiência. Tendo em conta o aumento de risco de erros de modelação ao trabalhar com estimadores de informação completa, não é fácil encontrar motivação que sustente a utilização dos *MIC*, embora seja um processo mais elegante do ponto de vista matemático. Existe, assim, uma ambígua supremacia do estimador de informação completa em relação ao estimador de informação limitada, uma vez que a superioridade dos *MIC* é válida no limite, mas pode não se verificar em termos de aplicações práticas.

Os programas computacionais usados na obtenção das estimativas foram desenvolvidos pela autora, recorrendo a *software* livre, nomeadamente ao programa *R* (disponível em <http://www.R-project.org>, ver *R* development Core Team, 2007) e a algumas das suas *packages* específicas. No entanto, a *package systemfit*, que está prevista para o tratamento computacional dos *SEM*, apenas contempla o caso particular em que os modelos são homocedásticos. Para além disso, não inclui a estimação por *GMM*, nem existe actualmente nenhuma outra *package* do *R* que contemple esse método, pelo menos devidamente testada e divulgada. Assim, apenas foi possível usar a *package systemfit* para calcular e comparar as estimativas de mínimos quadrados (obtidas por *MIL* e *MIC*). Para as comparações que envolveram *GMM*, foi necessário que a autora

desenvolvesse um conjunto de macros, que contemplasse o método adaptado à situação mais geral de heterocedasticidade, quer para a informação limitada quer para a informação completa.

Deste modo, os exemplos que se seguem, para além de ilustrar os processos de cálculo das estimativas e o desempenho dos estimadores, serviram adicionalmente para certificar o funcionamento das macros desenvolvidas pela autora no programa *R*, durante a realização deste trabalho. A confirmação dos resultados relativos à utilização do método dos mínimos quadrados foi feita comparando os resultados agora encontrados com os valores obtidos usando a *package systemfit*. Como era de esperar, os resultados foram semelhantes. Em relação ao programa utilizado para obter as estimativas por *GMM* não existem análises de dados publicadas que possibilitem essa comparação. No entanto, as conclusões foram semelhantes às que se obtiveram usando o método dos mínimos quadrados, o que leva a não duvidar no programa desenvolvido.

### 3.10.1 Modelo proposto por Kmenta (1997)

No primeiro exemplo estudou-se um modelo proposto por Kmenta (1997) com dois objectivos: o primeiro, de validar o programa desenvolvido no presente trabalho para obter estimativas dos parâmetros estruturais do modelo, de acordo com diversos métodos de estimação; o segundo, de confirmar os resultados teóricos estabelecidos em relação à eficiência relativa entre estimadores de informação limitada e de informação completa. Assim, para cumprir o primeiro objectivo, obtiveram-se estimativas dos parâmetros estruturais por duas vias distintas: usando directamente as funções disponíveis no *systemfit* ou aplicando o programa desenvolvido pela autora. A validação do programa foi conseguida, na medida em que se obtiveram resultados idênticos por ambos os processos.

*Exemplo 3.10.1.* O *SEM* tratado neste exemplo, bem como o correspondente conjunto de dados foram propostos por Kmenta (1997). Trata-se de um *SEM* composto por duas equações, referindo-se a primeira equação à "função procura" e a segunda equação à "função oferta", o qual pode ser definido pelo seguinte sistema:

$$\begin{cases} q &= \alpha_0 + p\alpha_1 + d\alpha_2 + e_1 \\ q &= \beta_0 + p\beta_1 + f\beta_2 + a\beta_3 + e_2 \end{cases} \quad (3.10.1)$$

O sistema contém as variáveis endógenas  $q$  e  $p$ :  $q$  representa o consumo com a alimentação *per capita* e  $p$  traduz a relação entre o preço da alimentação e o preço geral no consumidor. Para além dessas variáveis, o sistema inclui as variáveis exógenas  $d$ ,  $f$  e  $a$ :  $d$  refere-se ao rendimento disponível,  $f$  traduz a proporção do preço que é pago ao produtor (agrícola) no ano anterior, em relação ao preço pago pelo consumidor e  $a$  representa o tempo em anos.

Para apresentar o modelo na forma estrutural, referida em (2.2.2), basta fazer algumas transformações na forma anterior, seguidas de uma mudança de variáveis, o que conduz ao seguinte

$$\begin{cases} q &= \alpha_0 + p\alpha_1 + d\alpha_2 + e_1 \\ p &= -\frac{\beta_0}{\beta_1} + q\frac{1}{\beta_1} - f\frac{\beta_2}{\beta_1} - a\frac{\beta_3}{\beta_1} - \frac{e_2}{\beta_1} \end{cases} \Leftrightarrow \begin{cases} q &= \alpha_0 + p\alpha_1 + d\alpha_2 + e_1 \\ p &= \beta_0^* + q\beta_1^* + f\beta_2^* + a\beta_3^* + e_2^* \end{cases}, \quad (3.10.2)$$

onde se considerou

$$-\frac{\beta_i}{\beta_1} = \beta_i^*, i = 0, 2, 3, \frac{1}{\beta_1} = \beta_1^* \text{ e } -\frac{e_2}{\beta_1} = e_2^*.$$

Os dados utilizados correspondem a 20 realizações anuais das cinco variáveis já referidas e estão disponíveis no programa *R*. Este conjunto de dados foi construído por Kmenta para ilustrar o processo de estimação de um *SEM*. Os valores das variáveis exógenas, **d**, **f** e **a**, são baseados em dados reais, enquanto que os das variáveis endógenas, **q** e **p** foram gerados por simulação por Kmenta.

A partir de (3.10.2), pode obter-se o sistema na forma estrutural

$$\begin{cases} \mathbf{p}\alpha_1 - \mathbf{q} + \alpha_0 + \mathbf{d}\alpha_2 + \mathbf{e}_1 &= 0 \\ -\mathbf{p} + \mathbf{q}\beta_1^* + \beta_0^* + \mathbf{f}\beta_2^* + \mathbf{a}\beta_3^* + \mathbf{e}_2^* &= 0 \end{cases}. \quad (3.10.3)$$

Confirmou-se que se tratava de um sistema completo, na medida em que o número de equações coincide com o número de variáveis endógenas, duas neste caso.

Estudou-se depois a identificabilidade de cada equação, recorrendo à **Proposição 2.3.4** e tendo em conta que a matriz dos parâmetros estruturais do modelo da forma (3.10.3)

é:

$$\Delta = \begin{bmatrix} \alpha_1 & -1 \\ -1 & \beta_1^* \\ \alpha_0 & \beta_0^* \\ \alpha_2 & 0 \\ 0 & \beta_2^* \\ 0 & \beta_3^* \end{bmatrix}.$$

Assim, para a primeira equação, obteve-se a matriz, de acordo com (2.3.6),

$$\mathbf{R}_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

formada pelos coeficientes das variáveis nas restrições de exclusão da primeira equação - cada linha está associada a uma restrição de exclusão - e verificou-se que  $\text{rank}(\mathbf{R}_1\Delta) = \text{rank} \begin{bmatrix} 0 & \beta_2^* \\ 0 & \beta_3^* \end{bmatrix} = 1$ , enquanto que  $\text{rank}(\mathbf{R}_1) = 2 > 1$ . Concluiu-se, assim, que a primeira equação é sobreidentificada.

Quanto à segunda equação, uma vez que

$$\mathbf{R}_2 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix},$$

vem que  $\text{rank}(\mathbf{R}_2\Delta) = \text{rank} \begin{bmatrix} \alpha_2 & 0 \end{bmatrix} = 1 = \text{rank}(\mathbf{R}_2)$ , pelo que a segunda equação é exactamente identificada.

Tendo-se concluído que todas as equações do modelo são identificadas passou-se à estimação dos parâmetros do modelo, recorrendo primeiro à *package systemfit*. Ao usar directamente essa *package*, foram introduzidas as equações de acordo com a forma (3.10.1), de modo que o parâmetro a estimar é o vector

$$\begin{bmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \beta_0 & \beta_1 & \beta_2 & \beta_3 \end{bmatrix}'.$$

Os resultados estão organizados na Tabela 3.2. Note-se que, em relação à primeira equação, as estimativas por *2SLS* e por *3SLS* coincidiram para todos os parâmetros. Em relação à segunda equação, isso não aconteceu, embora as estimativas por ambos os métodos sejam próximas. Esta conclusão não colide com a propriedade P3SLS.4, uma vez que nem todas as equações do sistema são exactamente identificadas.

equação	parâmetro	<i>2SLS</i> ( <i>systemfit</i> )	<i>3SLS</i> ( <i>systemfit</i> )
1	$\alpha_0$	94.63	94.63
1	$\alpha_1$	-0.24	-0.24
1	$\alpha_2$	0.31	0.31
2	$\beta_0$	49.53	52.20
2	$\beta_1$	0.24	0.23
2	$\beta_2$	0.26	0.23
2	$\beta_3$	0.25	0.36

Tabela 3.2: Estimativas 2SLS e 3SLS obtidas com systemfit.

equação	parâmetro	<i>2SLS</i>	<i>3SLS</i>
1	$\alpha_0$	94.63	94.63
1	$\alpha_1$	-0.24	-0.24
1	$\alpha_2$	0.31	0.31
2	$\beta_0$	49.59	51.99
2	$\beta_1$	0.24	0.23
2	$\beta_2$	0.25	0.23
2	$\beta_3$	0.25	0.24

Tabela 3.3: Estimativas 2SLS e 3SLS obtidas com o programa desenvolvido pela autora.

De seguida, aplicou-se o programa desenvolvido neste trabalho para obter estimativas do mesmo vector de parâmetros. As estimativas obtidas, quer por *2SLS* quer por *3SLS*, organizaram-se na Tabela 3.3, com os resultados correspondentes aos da Tabela 3.2.

Da comparação entre as Tabelas 3.2 e 3.3, pode concluir-se que as estimativas obtidas pelos dois processos de cálculo foram muito semelhantes. Em particular, foram coincidentes em relação a todos os parâmetros da primeira equação e a alguns da segunda equação, não se registando diferenças significativas nas restantes estimativas. Esta constatação contribuiu para validar o programa desenvolvido.

Posteriormente, no estudo da eficiência relativa dos estimadores de informação limitada e de informação completa, baseado no critério da variância generalizada, calcularam-se as diferenças entre os determinantes das matrizes de covariâncias estimadas por informação limitada e por informação completa.

Deste modo, para o método dos mínimos quadrados, obteve-se o valor

$$\left| \widehat{VA} \left[ \hat{\delta}_{2SLS} \right] \right| - \left| \widehat{VA} \left[ \hat{\delta}_{3SLS} \right] \right| = -2.27 \times 10^{-10},$$

que é muito próximo de zero, embora negativo.

Para aplicar o método dos momentos generalizado, como se referiu, não foi possível

equação	parâmetro	GMMIL	GMMIC
1	$\alpha_0$	95.68	95.68
1	$\alpha_1$	-0.24	-0.24
1	$\alpha_2$	0.30	0.30
2	$\beta_0$	49.59	53.22
2	$\beta_1$	0.24	0.22
2	$\beta_2$	0.25	0.23
2	$\beta_3$	0.25	0.33

Tabela 3.4: Estimativas GMMIL e GMMIC obtidas com o programa desenvolvido pela autora.

utilizar nenhuma *package* do *R*. Assim, obtiveram-se estimativas para os parâmetros, cujos valores constam da Tabela 3.4.

Relativamente à diferença dos determinantes, para o Método dos Momentos Generalizado, obteve-se o valor

$$\left| \widehat{VA} \left[ \hat{\delta}_{GMMIL} \right] \right| - \left| \widehat{VA} \left[ \hat{\delta}_{GMMIC} \right] \right| = 3.71 \times 10^{-11},$$

também muito próximo de zero, mas desta vez positivo. Observando a ordem de grandeza das diferenças e o facto de estar demonstrado que os *MIC* são assintoticamente superiores, pode considerar-se se essas diferenças não serão devidas sobretudo a questões numéricas. Este exemplo mostra-nos que, para esta dimensão, não é possível afirmar que os *MIC* são superiores aos *MIL*.  $\diamond$

### 3.10.2 Modelo proposto por Judge *et al.* (1988) - simulação de dados

O segundo exemplo desta secção consiste no estudo de um modelo proposto por Judge *et al.* (1988). Este exemplo envolve dados simulados pela autora e foi usado para apresentar detalhadamente o processo de simulação das amostras de acordo com um *SEM*, o que constituiu uma componente de trabalho significativa. Na realidade o modelo é complexo e a simulação de amostras não é trivial.

Para gerar as amostras de acordo com o *SEM*, utilizou-se o modelo econométrico apresentado em Judge *et al.* (1988), com os respectivos parâmetros conhecidos e as



observações das variáveis exógenas. O exemplo publicado continha vinte observações do modelo, cujos valores das variáveis endógenas tinham sido obtidos por simulação, a partir da distribuição normal dos erros. Tendo em vista fazer um estudo mais detalhado da estimação, nomeadamente, investigando o desempenho dos estimadores para diferentes dimensões amostrais e diferentes distribuições dos erros, decidiu-se simular amostras do modelo, mantendo as matrizes dos coeficientes estruturais  $\mathbf{\Gamma}$  e  $\mathbf{B}$  (conhecidas), e os valores das variáveis exógenas,  $\mathbf{X}$ , em dimensões amostrais em que tal era possível. Assim, as observações das variáveis endógenas foram obtidas por simulação dos erros, de acordo com quatro distribuições multivariadas e após a representação do modelo na forma reduzida.

Os resultados obtidos na estimação das matrizes de covariâncias dos estimadores foram posteriormente comparados através das diferença entre os determinantes das correspondentes matrizes estimadas, tendo-se verificado que a diferença foi quase nula em todas as situações especificadas.

*Exemplo 3.10.2.* Considere-se um *SEM* divulgado em Judge *et al.* (1988), formado por três equações, com três variáveis endógenas,  $y_1, y_2, y_3$ , e cinco variáveis exógenas,  $x_1, x_2, x_3, x_4, x_5$ , definido do seguinte modo:

$$\begin{cases} y_1 = & y_2\gamma_{21} + y_3\gamma_{31} + x_1\beta_{11} & + e_1 \\ y_2 = y_1\gamma_{12} & + x_1\beta_{12} + x_2\beta_{22} + x_3\beta_{32} + x_4\beta_{42} & + e_2 \\ y_3 = & y_2\gamma_{23} & + x_1\beta_{13} + x_2\beta_{23} & + x_5\beta_{53} + e_3 \end{cases} .$$

Para uma amostra genérica de  $T$  observações do modelo, a forma estrutural, correspondente a  $\mathbf{Y}\mathbf{\Gamma} + \mathbf{X}\mathbf{B} + \mathbf{E} = \mathbf{0}$ , é a seguinte:

$$\begin{cases} -\mathbf{Y}_1 & + \mathbf{Y}_2\gamma_{21} + \mathbf{Y}_3\gamma_{31} + \mathbf{X}_1\beta_{11} & + \mathbf{e}_1 & = \mathbf{0} \\ \mathbf{Y}_1\gamma_{12} - \mathbf{Y}_2 & + \mathbf{X}_1\beta_{12} + \mathbf{X}_2\beta_{22} + \mathbf{X}_3\beta_{32} + \mathbf{X}_4\beta_{42} & + \mathbf{e}_2 & = \mathbf{0} \\ \mathbf{Y}_2\gamma_{23} - \mathbf{Y}_3 & + \mathbf{X}_1\beta_{13} + \mathbf{X}_2\beta_{23} & + \mathbf{X}_5\beta_{53} + \mathbf{e}_3 & = \mathbf{0} \end{cases} ,$$

com matrizes de parâmetros

$$\mathbf{\Gamma}_{(3 \times 3)} = \begin{bmatrix} -1 & \gamma_{12} & 0 \\ \gamma_{21} & -1 & \gamma_{23} \\ \gamma_{31} & 0 & -1 \end{bmatrix}, \mathbf{B}_{(5 \times 3)} = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ 0 & \beta_{22} & \beta_{23} \\ 0 & \beta_{32} & 0 \\ 0 & \beta_{42} & 0 \\ 0 & 0 & \beta_{53} \end{bmatrix}$$

e matriz de erros

$$\mathbf{E}_{(T \times 3)} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{e}'_1 \\ \vdots \\ \mathbf{e}'_T \end{bmatrix}.$$

Antes de simular a amostra, verificou-se que o sistema é completo, uma vez que o número de variáveis endógenas é igual ao número de equações do modelo, neste caso é igual a três.

De seguida, estudou-se a identificabilidade de cada equação, tendo-se verificado que a primeira e a terceira equações são sobreidentificadas, enquanto que a segunda equação é exactamente identificada. Concluiu-se assim que todas as equações do modelo são identificadas. Para isso, considerou-se a matriz dos parâmetros estruturais do modelo,

$$\Delta_{((3+5) \times 3)} = \begin{bmatrix} \mathbf{\Gamma} \\ \mathbf{B} \end{bmatrix} = \begin{bmatrix} -1 & \gamma_{12} & 0 \\ \gamma_{21} & -1 & \gamma_{23} \\ \gamma_{31} & 0 & -1 \\ \beta_{11} & \beta_{12} & \beta_{13} \\ 0 & \beta_{22} & \beta_{23} \\ 0 & \beta_{32} & 0 \\ 0 & \beta_{42} & 0 \\ 0 & 0 & \beta_{53} \end{bmatrix},$$

e, para cada equação, definiu-se a matriz dos coeficientes das restrições de exclusão da equação  $i$  na forma estrutural,  $\mathbf{R}_i$ , com base na qual se procedeu à discussão da identificabilidade, de acordo com a **Proposição 2.3.4**.

Assim, para a primeira equação, como

$$\begin{aligned} \mathbf{R}_1 \Delta &= \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \Delta \\ &= \begin{bmatrix} 0 & \beta_{21} & \beta_{23} \\ 0 & \beta_{32} & 0 \\ 0 & \beta_{42} & 0 \\ 0 & 0 & \beta_{53} \end{bmatrix}, \end{aligned}$$

vem que  $\text{rank}(\mathbf{R}_1\Delta) = 2$  e  $\text{rank}(\mathbf{R}_1) = 4 > 2$ , logo esta equação é sobreidentificada.

Para a segunda equação, tem-se que

$$\begin{aligned}\mathbf{R}_2\Delta &= \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \Delta \\ &= \begin{bmatrix} \gamma_{31} & 0 & -1 \\ 0 & 0 & \beta_{53} \end{bmatrix},\end{aligned}$$

pelo que  $\text{rank}(\mathbf{R}_2\Delta) = 2$  e  $\text{rank}(\mathbf{R}_2) = 2$ , donde se concluiu que a segunda equação é exactamente identificada.

Na terceira equação, uma vez que

$$\begin{aligned}\mathbf{R}_3\Delta &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \Delta \\ &= \begin{bmatrix} -1 & \gamma_{12} & 0 \\ 0 & \beta_{32} & 0 \\ 0 & \beta_{42} & 0 \end{bmatrix},\end{aligned}$$

vem que  $\text{rank}(\mathbf{R}_3\Delta) = 2$  e  $\text{rank}(\mathbf{R}_3) = 3 > 2$ , o que permitiu concluir que a equação é sobreidentificada.

Passou-se de seguida à simulação de amostras de dimensão  $T$ , com base nas quais se procedeu à estimação dos parâmetros. Todos os cálculos foram efectuados com o programa *R*. Para proceder à simulação das amostras, seguiram-se as etapas que se passam a descrever.

Fixaram-se previamente as matrizes dos coeficientes  $\mathbf{\Gamma}$  e  $\mathbf{B}$  com os valores usados em Judge *et al.* (1988), isto é,

$$\mathbf{\Gamma} = \begin{bmatrix} -1 & 0.2 & 0 \\ -10 & -1 & 2 \\ 2.5 & 0 & -1 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} -60 & 40 & -10 \\ 0 & -4 & 80 \\ 0 & -6 & 0 \\ 0 & 1.5 & 0 \\ 0 & 0 & 5 \end{bmatrix}.$$

Tendo em conta a existência de vinte valores das variáveis exógenas, no estudo de simulação efectuado, gerou-se uma amostra de dimensão  $T = 20$ . Utilizaram-se os vinte

valores das variáveis exógenas usadas por Judge *et al.* (1988),

$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \mathbf{X}_3 & \mathbf{X}_4 & \mathbf{X}_5 \end{bmatrix}_{(20 \times 5)}$ . Posteriormente gerou-se ainda uma outra amostra, de dimensão  $T = 120$ , para investigar o efeito do aumento da dimensão da amostra no estudo da eficiência relativa empírica dos estimadores.

A simulação das variáveis endógenas foi efectuada a partir da geração dos erros do modelo. Para cada dimensão amostral, foram consideradas quatro distribuições dos erros  $\mathbf{e}_t = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \end{bmatrix}_t'$ :

D.1 Normal trivariada  $N_3(\mathbf{0}, \mathbf{\Sigma})$ , com

$$\mathbf{\Sigma} = \begin{bmatrix} 227.6 & 9 & -57 \\ 9 & 0.7 & -2 \\ -57 & -2 & 15.8 \end{bmatrix}, \quad (3.10.4)$$

usada em Judge *et al.* (1988).

D.2 Normal trivariada  $N_3(\mathbf{0}, \mathbf{\Sigma}^*)$ , com  $\mathbf{\Sigma}^* = \begin{bmatrix} 227.6 & 4 & -21 \\ 4 & 0.7 & -0.7 \\ -21 & -0.7 & 15.8 \end{bmatrix}$ . A matriz  $\mathbf{\Sigma}^*$

difere de  $\mathbf{\Sigma}$  apenas nas covariâncias, pois verificou-se que correspondiam a valores elevados de coeficientes de correlação entre alguns erros. Na realidade, a matriz de

correlações correspondente a  $\mathbf{\Sigma}$  é  $\mathbf{CO} = \begin{bmatrix} 1 & 0.73 & -0.95 \\ 0.73 & 1 & -0.58 \\ -0.95 & -0.58 & 1 \end{bmatrix}$ . A existência de

erros altamente correlacionados podia ser enganadora ao estabelecer conclusões com base na variância generalizada. Assim, simularam-se erros de acordo com a matriz de

correlações  $\mathbf{CO}^* = \begin{bmatrix} 1 & 0.36 & -0.22 \\ 0.36 & 1 & -0.14 \\ -0.22 & -0.14 & 1 \end{bmatrix}$ . Foi com base em  $\mathbf{CO}^*$  que se decidiu a escolha de  $\mathbf{\Sigma}^*$ .

D.3 Dados transformados a partir de distribuições marginais uniformes em  $(-1, 1)$  e matriz de covariâncias  $\mathbf{\Sigma}$ .

D.4 Dados transformados a partir de distribuições marginais logísticas e matriz de covariâncias  $\mathbf{\Sigma}$ .

A simulação das amostras com dimensão  $T = 20$  não envolveu nenhum procedimento adicional. Para estudar o desempenho dos estimadores, com as amostras de dimensão mais elevada ( $T = 120$ ), foi necessário gerar também valores adicionais para as

variáveis exógenas. Procurando manter características análogas às das 20 observações publicadas de  $x_1, x_2, x_3, x_4, x_5$ , procedeu-se do seguinte modo:

a) Para cada uma das variáveis  $x_2, x_3, x_4$  geraram-se valores de uma distribuição uniforme com medidas amostrais idênticas às das observações das variáveis exógenas no exemplo publicado.

b) Para  $x_1$  e  $x_5$ , procedeu-se de modo diferente:  $x_1$ , porque apenas tomava valores unitários; e  $x_5$  porque tomava valores inteiros, pelo que as observações foram geradas por uma uniforme discreta.

De acordo com as alíneas a) e b) anteriores, simularam-se 100 observações independentes de variáveis exógenas, que se juntaram às 20 já existentes para prosseguir com a simulação da amostra de  $T = 120$ .

Para cada distribuição dos erros considerada, verificou-se que as conclusões sobre a eficiência relativa empírica dos estimadores eram idênticas para ambas as dimensões amostrais, mostrando assim que não havia ganho de eficiência com o aumento da dimensão da amostra. Na Tabela 3.6 apresentam-se os resultados para a dimensão  $T = 20$ , quando os erros seguem distribuição normal - D.1 e D.2. Para as restantes distribuições, apresentam-se apenas os resultados relativos a  $T = 120$ .

Para cada uma das situações amostrais descritas, a partir dos valores de

$\mathbf{E} = \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_T \end{bmatrix}'_{(T \times 3)}$ , obteve-se a matriz dos erros da forma reduzida,  $\mathbf{V} = -\mathbf{E}\mathbf{\Gamma}^{-1}$ ; depois, calculou-se a matriz  $\mathbf{\Pi} = -\mathbf{B}\mathbf{\Gamma}^{-1}$ , e, por fim, obtiveram-se os valores das variáveis endógenas,  $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3]_{(T \times 3)}$ , substituindo as matrizes obtidas anteriormente na forma reduzida,  $\mathbf{Y} = \mathbf{X}\mathbf{\Pi} + \mathbf{V}$ .

Deste modo, resultaram os valores  $\mathbf{y}_i$  e gerou-se cada observação da forma

$\begin{pmatrix} y_1 & y_2 & y_3 & x_1 & \cdots & x_5 \end{pmatrix}$ , onde  $y_i$  são os valores das variáveis endógenas e  $x_j$  são os valores das variáveis exógenas.

Para passar à estimação dos parâmetros, supondo  $\mathbf{\Pi}$  desconhecida, geraram-se  $T$  observações da forma imediatamente antes referida e definiu-se a matriz de dimensão  $(T \times 8)$ ,

$$\begin{bmatrix} \mathbf{Y} & \mathbf{X} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \mathbf{y}_3 & \mathbf{x}_1 & \cdots & \mathbf{x}_5 \end{bmatrix}.$$

Note-se que cada  $\mathbf{y}_i$  e cada  $\mathbf{x}_j$  ( $i = 1, 2, 3, j = 1, \dots, 5$ ) são vectores  $T$ -dimensionais.

Deste modo foi possível dispor de uma amostra simulada, de dimensão  $T$ , para cada

uma das situações amostrais consideradas.

Para cada amostra, estimaram-se os parâmetros do modelo usando os dois pares de estimadores:  $\hat{\delta}_{2SLS}$  e  $\hat{\delta}_{3SLS}$ , e  $\hat{\delta}_{GMMIL}$  e  $\hat{\delta}_{GMMIC}$ .

Para qualquer dos métodos usados, houve necessidade de estimar  $\Omega$  previamente. No caso do estimador  $\hat{\delta}_{3SLS}$ , definido por (3.4.2), a matriz  $\Omega = \Sigma \otimes \mathbf{I}$  é uma matriz por blocos, em que os blocos são diagonais com elementos iguais dentro de cada bloco. A igualdade desses elementos cria condições que facilitam a estimação de  $\Omega$ .

Ao trabalhar com o estimador  $GMM$ , não é pressuposta a homocedasticidade e, por isso, foi necessário estimar cada um dos elementos dos blocos - pois  $\Omega$  é formada por blocos diagonais com elementos diferentes dentro de cada bloco. Tradicionalmente, a estimação de cada bloco é conseguida aplicando o estimador de White, definido por

$$\hat{\mathbf{W}}_{ij} = \mathbf{X}'\hat{\Omega}_{ij}\mathbf{X}.$$

A matriz  $\hat{\Omega}_{ij}$  é formada por elementos diagonais que, segundo Greene (2003), são os produtos dos resíduos obtidos por  $2SLS$ , isto é, são da forma

$$(\hat{e}_{it,2SLS})(\hat{e}_{jt,2SLS}) = \left(y_{it} - \mathbf{Z}_{it}\hat{\delta}_{i,2SLS}\right)\left(y_{jt} - \mathbf{Z}_{jt}\hat{\delta}_{j,2SLS}\right).$$

A matriz  $\hat{\Omega}_i$  é formada por elementos diagonais que são os quadrados dos resíduos correspondentes, isto é, tais que

$$(\hat{e}_{it,2SLS})^2 = \left(y_{it} - \mathbf{Z}_{it}\hat{\delta}_{i,2SLS}\right)^2.$$

Na Tabela 3.5 organizaram-se as estimativas obtidas para os parâmetros estruturais, para o caso dos erros com distribuição D.1 ( $N_3(\mathbf{0}, \Sigma)$ ) e para uma amostra de dimensão  $T = 20$ .

Para comparar as estimativas quanto à eficiência, a análise dos resultados baseou-se no critério da variância generalizada. Os resultados obtidos foram comparados através das diferenças entre os determinantes das estimativas das matrizes de covariâncias dos estimadores de informação limitada e completa. Na Tabela 3.6 mostram-se os valores em causa para as diferentes amostras simuladas e para cada um dos processos de estimação em estudo.

valores usados na simulação		estimativas de mínimos quadrados		estimativas por <i>GMM</i>	
		<i>2SLS</i>	<i>3SLS</i>	<i>GMMIL</i>	<i>GMMIC</i>
$\Gamma$	$\gamma_{12} = 0.2$	0.203	0.200	0.203	0.203
$\Gamma$	$\gamma_{21} = -10$	-9.194	-9.360	-8.698	-8.601
$\Gamma$	$\gamma_{23} = 2$	2.865	1.693	2.36	1.865
$\Gamma$	$\gamma_{31} = 2.5$	2.375	2.402	2.298	2.283
$B$	$\beta_{11} = -60$	-69.702	-67.699	-75.838	-77.164
$B$	$\beta_{12} = 40$	40.519	40.459	40.519	40.113
$B$	$\beta_{13} = -10$	-19.863	-5.583	-13.609	-7.708
$B$	$\beta_{22} = -4$	-4.425	-4.379	-4.425	-4.376
$B$	$\beta_{23} = 80$	64.642	85.587	73.748	82.570
$B$	$\beta_{32} = -6$	-5.242	-5.453	-5.242	-5.599
$B$	$\beta_{42} = 1.5$	1.338	1.464	1.338	1.417
$B$	$\beta_{53} = 5$	3.779	5.357	4.460	5.109
$\Sigma$	$\sigma_{11} = 227.6$	128.546	131.209	122.387	121.518
$\Sigma$	$\sigma_{12} = 9$	4.477	6.175	4.138	5.155
$\Sigma$	$\sigma_{13} = -57$	-30.472	-32.879	-30.220	-30.959
$\Sigma$	$\sigma_{22} = 0.7$	1.012	1.068	1.012	1.057
$\Sigma$	$\sigma_{23} = -2$	-1.405	-0.862	-0.978	-0.746
$\Sigma$	$\sigma_{33} = 15.8$	11.133	10.766	10.154	10.264

Tabela 3.5: Estimativas dos parâmetros estruturais obtidas com T=20 observações, usando variáveis exógenas e matrizes  $\Gamma$ ,  $B$  e  $\Sigma$  de acordo com Judge e variáveis endógenas simuladas com erros  $N(0, \Sigma)$ .

Distribuição	$\widehat{VA}(\hat{\delta}_{2SLS})$	-	$\widehat{VA}(\hat{\delta}_{3SLS})$	$\widehat{VA}(\hat{\delta}_{GMMIL})$	-	$\widehat{VA}(\hat{\delta}_{GMMIC})$
$N(\mathbf{0}, \Sigma)$ - dim 20	$-1.01 \times 10^{-20}$			$1.45 \times 10^{-20}$		
$N(\mathbf{0}, \Sigma)$ - dim 120	$-4.72 \times 10^{-37}$			$4.07 \times 10^{-33}$		
$N(\mathbf{0}, \Sigma^*)$ - dim 20	$-1.74 \times 10^{-16}$			$1.80 \times 10^{-18}$		
$U(-1, 1)$ - dim 120	$3.03 \times 10^{-43}$			$3.00 \times 10^{-38}$		
Logística- dim 120	$-1.79 \times 10^{-32}$			$3.71 \times 10^{-27}$		

Tabela 3.6: Valores das diferenças entre os determinantes das estimativas das matrizes de covariâncias dos estimadores de informação limitada e completa para as diferentes amostras.

Verificou-se que a diferença entre os determinantes é quase nula em todas as situações consideradas, podendo dizer-se que é nula, em termos de efeitos práticos. De facto, as diferenças entre determinantes são tão pequenas que podem ser devidas apenas a questões numéricas. Em todo o caso, ao usar o método dos momentos generalizado, essas diferenças são sempre positivas, o que aponta no sentido do resultado verificado teoricamente.

Por outro lado, os resultados mostram que a mesma conclusão é válida independentemente da dimensão da amostra usada, pois, supondo a distribuição normal, não houve alterações significativas (do ponto de vista prático) quando se passou de uma amostra de dimensão 20 para uma amostra de dimensão 120. Admitindo as distribuições uniforme ou a logística, os estimadores deram estimativas ainda mais próximas do que no caso da distribuição normal.

Embora tenha sido verificado teoricamente que só pode haver vantagem em usar métodos de informação completa sobre os métodos de informação limitada, nomeadamente *3SLS* sobre *2SLS* e *GMMIC* sobre *GMMIL*, os resultados computacionais não evidenciam essa vantagem.

Conclui-se assim que, para dimensão finita, não há diferença significativa entre usar *MIL* ou *MIC*. Importa referir que esta conclusão não parece ser influenciada pelo método de estimação de base nem pela dimensão da amostra.

Deste modo, tendo em conta que para utilizar os *MIC* se torna necessário manipular matrizes de dimensão elevada, pode não se justificar a informação completa, por não conduzir a vantagens relevantes nos resultados finais. Repare-se, a título de exemplo, que no *SEM* estudado neste exemplo, enquanto que com um *MIL* os cálculos envolvem matrizes de dados de dimensão  $(20 \times 8)$  [ou  $(120 \times 8)$ ], num *MIC* as matrizes de dados usadas são do tipo  $(60 \times 8)$  [ou  $(360 \times 8)$ ].

De acordo com o **Teorema 3.9.5**, sabe-se que, assintoticamente, o estimador *GMMIC* é tão eficiente, ou mais, do que o estimador *GMMIL*. Mas esta conclusão é válida em termos assintóticos e não em dimensão finita.

Resumindo, este exemplo sugere os seguintes comentários:



- (i) Em dimensão finita, geralmente, o estimador  $2SLS$  apresenta menor variância generalizada do que a do  $3SLS$ . Note-se que isto não contradiz os resultados demonstrados.
- (ii) Não há ganho significativo com o aumento da dimensão da amostra. Quando se trabalha com erros normais, a dimensão 20 permitiu obter as mesmas conclusões que a dimensão 120.
- (iii) Para amostras de pequena dimensão é quase indiferente usar  $2SLS$  ou  $3SLS$ , assim como  $GMMIL$  ou  $GMMIC$ .
- (iv) Mesmo nos casos considerados, em que os erros não tinham distribuição normal (uniforme ou logística), as conclusões anteriores continuam a observar-se.
- (v) Os resultados obtidos confirmam os comentários apresentados no exemplo anterior, nomeadamente que, para dimensões finitas e em termos de eficiência, não há grande diferença entre usar  $MIL$  ou  $MIC$ . Esta conclusão parece ser válida independentemente da distribuição dos dados, do método de estimação de base, ou da dimensão da amostra.

◇

## Capítulo 4

# Introdução à estimação robusta

Para estudar a estimação robusta nos *SEM*, convém ter presentes os princípios da estatística robusta e algumas das suas noções e técnicas básicas.

A estatística robusta tem como objectivo o desenvolvimento e a aplicação de métodos que sejam capazes de apresentar um bom desempenho, quando a realidade se afasta um pouco dos pressupostos assumidos nos modelos estatísticos.

Grande parte dos métodos estatísticos de maior divulgação, nomeadamente os que mais se utilizam para realizar testes de hipóteses, determinar intervalos de confiança, estimar coeficientes de regressão, entre outros, exigem a verificação de um conjunto de pressupostos, como, por exemplo, a normalidade dos dados, a existência de observações provenientes de amostras aleatórias, isto é, formadas por variáveis independentes e identicamente distribuídas (*i.i.d.*) ou a homocedasticidade. Caso não se verifique algum dos pressupostos de um modelo estatístico, não há a garantia de que se mantenham as boas propriedades que motivam a utilização desses métodos estatísticos. Isto é, os referidos métodos que, em geral, são óptimos quando os pressupostos do modelo são verdadeiros, perante a falha de tais pressupostos, podem conduzir a resultados desastrosos, pondo em causa a sua própria utilização.

Quando um método estatístico é robusto, os resultados que produz não devem mostrar grandes alterações perante ligeiros desvios dos pressupostos assumidos. É neste sentido que se interpreta a noção de robustez estatística; ou seja, considera-se que um método estatístico é robusto quando é pouco sensível a ligeiros afastamentos dos pressupostos assumidos para o modelo.

Ao considerar a existência de observações atípicas como um desvio do modelo assumido, pode encarar-se essa questão como um ponto de interesse da estatística robusta. De facto, os métodos robustos podem constituir um importante instrumento de detecção de observações atípicas. Para além disso, ao nível da análise estatística, os métodos robustos permitem controlar e/ou atenuar os efeitos dessas observações.

Apesar das vantagens referidas, os métodos robustos ainda são pouco utilizados. Para além de serem técnicas desenvolvidas mais recentemente, a sua menor utilização também se deve ao facto de existir uma grande variedade de métodos robustos, não sendo fácil para o utilizador escolher o que mais lhe convém. Por outro lado, estes métodos não estão ainda suficientemente divulgados; veja-se, por exemplo, que é um tópico que só agora começa a aparecer nos livros básicos de estatística. Acresce ainda o facto de, em geral, exigirem uma componente computacional significativa, para o cálculo das estimativas.

Na apresentação dos conceitos gerais que se fará, segue-se de perto a orientação de Pires e Branco (2007), que constitui actualmente a única publicação introdutória escrita em português sobre estatística robusta.

## 4.1 Conceitos gerais

Nesta secção pretende-se apresentar os principais conceitos em que se baseia a estatística robusta.

Uma das noções mais simples e intuitiva é a de curva de sensibilidade de um estimador. Trata-se de uma medida amostral que evidencia com facilidade o efeito que uma observação adicional pode ter sobre uma estimativa.

**Definição 4.1.1.** Considere-se uma realização  $(x_1, \dots, x_{n-1})$  de uma amostra aleatória  $(X_1, \dots, X_{n-1})$ , retirada de uma população com distribuição dependente de um parâmetro  $\theta \in \Theta$ . Seja  $T = T(X_1, \dots, X_n)$  um estimador de  $\theta$  e considere-se uma perturbação da amostra observada, pela inclusão de uma nova observação  $x$ , ou seja, passando a ter-se  $(x_1, \dots, x_{n-1}, x)$ . A **curva de sensibilidade** do estimador  $T$  para essa amostra  $(x_1, \dots, x_{n-1}, x)$  é uma função de  $x$ , definida por:

$$SC_n(x, T) = n [T(x_1, \dots, x_{n-1}, x) - T(x_1, \dots, x_{n-1})].$$

□

A ideia consiste em comparar os valores das estimativas obtidas por  $T$  para a amostra original e para a amostra perturbada. Pode dizer-se que é desejável que um estimador tenha curva de sensibilidade limitada. Ao multiplicar por  $n$  a diferença das estimativas, obtidas com a amostra perturbada e com a amostra original, pretende-se evitar o efeito da dimensão da amostra, como ilustra o exemplo que se segue.

*Exemplo 4.1.1.* Considere-se a média amostral como estimador da média de uma população e determine-se a curva de sensibilidade da média amostral. Neste caso, tem-se o estimador  $T = \bar{X}_n$ , evidenciando a dimensão da amostra, com  $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$ . A curva de sensibilidade vem

$$\begin{aligned}
 SC_n(x, T) &= n [T(x_1, \dots, x_{n-1}, x) - T(x_1, \dots, x_{n-1})] \\
 &= n (\bar{x}_n - \bar{x}_{n-1}) \\
 &= n \left[ \frac{1}{n} \left( \sum_{i=1}^{n-1} x_i + x \right) - \frac{1}{n-1} \sum_{i=1}^{n-1} x_i \right] \\
 &= \frac{n}{n(n-1)} \left( (n-1) \sum_{i=1}^{n-1} x_i + (n-1)x - n \sum_{i=1}^{n-1} x_i \right) \\
 &= \frac{n}{n(n-1)} \left( (n-1)x - \sum_{i=1}^{n-1} x_i \right) \\
 &= (x - \bar{x}_{n-1}).
 \end{aligned}$$

Pode verificar-se que esta função de  $x$  não é limitada, o que indica que um único valor  $x$  pode conduzir a um valor enorme da curva de sensibilidade de  $\bar{X}_n$ . Isto significa que  $\bar{X}_n$  é muito sensível a cada observação  $x$ , a qual pode distorcer o valor da estimativa obtida com a média amostral. ◇

Nem todas as medidas de localização têm a desvantagem referida: por exemplo, a mediana ou a média aparada apresentam curva de sensibilidade limitada. A curva de sensibilidade também pode ser calculada para medidas de dispersão e, tal como acontece com a média amostral, as medidas mais vulgarizadas também têm curva de sensibilidade ilimitada. De facto, ao calcular as curvas de sensibilidade da variância e do desvio padrão, obtêm-se funções quadráticas de  $x$ ; pelo que o efeito de uma

observação é agravado pelo expoente. Deste modo, a inclusão do valor  $x$  na amostra, pode ter efeitos mais acentuados sobre a estimativa do que acontecia no caso da média amostral. No entanto, nem todas as medidas de dispersão usuais possuem curva de sensibilidade ilimitada. Por exemplo, o desvio absoluto mediano ou  $MAD$  (de *Median Absolute Deviation*), que é definido pela mediana dos desvios absolutos das observações em relação à sua mediana, tem curva de sensibilidade limitada.

Para além da curva de sensibilidade, que é uma medida empírica da sensibilidade do estimador, passam a apresentar-se outras noções essenciais para a sistematização da estatística robusta.

Em primeiro lugar, no desenvolvimento teórico da robustez, houve vantagem em interpretar um estimador como um funcional estatístico. Para simplificar a apresentação dos conceitos, considera-se que tanto a população como os parâmetros são unidimensionais. A generalização a dimensões superiores é feita com naturalidade recorrendo às necessárias adaptações.

Seja  $(X_1, \dots, X_n)$  uma amostra aleatória de dimensão  $n$ , retirada de uma população  $X$ . Denote-se por  $F_\theta$  uma família de distribuições definidas sobre  $\mathbb{R}$ , dependentes de um parâmetro  $\theta$ , que toma valores no espaço de parâmetro  $\Theta \subset \mathbb{R}$ ; e designe-se por  $f_\theta$  a função de probabilidade, correspondente a  $F_\theta$ , ou a função densidade de probabilidade, conforme o caso. Represente-se ainda por  $\mathcal{F}$  o conjunto de todas as funções de distribuição de probabilidade que podem ser definidas sobre o espaço amostral. Pretende-se estimar  $\theta$  a partir de uma amostra  $(x_1, \dots, x_n)$ , correspondente a uma concretização de  $(X_1, \dots, X_n)$ .

Neste contexto, define-se um funcional estatístico  $T$ , como sendo uma aplicação que, a cada elemento de um espaço de funções de distribuição, faz corresponder um elemento de  $\mathbb{R}$ . Dito doutro modo, considerando que  $\mathcal{D}(T) \subset \mathcal{F}$  representa o domínio de  $T$ , o funcional estatístico  $T$  é uma aplicação

$$\begin{aligned} T : \mathcal{D}(T) &\rightarrow \mathbb{R} \\ F &\rightarrow T(F). \end{aligned}$$

Grande parte dos estimadores correntemente utilizados podem ser interpretados como

funcionais estatísticos, tendo em conta a seguinte condição: Seja  $G_n$  a função distribuição empírica da amostra; um estimador  $T$  de  $\theta$  é um funcional estatístico se pode sempre ser representado como um funcional de  $G_n$

$$T = T(X_1, \dots, X_n) = T(G_n), \forall n, \forall G_n.$$

Note-se que  $T$  depende de  $n$  apenas através de  $G_n$ . Nestas condições, a sucessão  $(T_n)_{n \in \mathbb{N}}$ , que fica definida para diferentes valores de  $n$  por  $T_n = T(X_1, \dots, X_n)$ , pode substituir-se por um funcional independente de  $n$ .

Por vezes, um estimador não tem representação funcional, no sentido em que  $T_n$  não se pode exprimir como  $T(G_n)$ . Mas, para a maior parte dos estimadores conhecidos, existe um funcional  $T$  para o qual se verifica que  $T_n(X_1, \dots, X_n) - T(G_n) = O(1/n)$ , onde  $O(1/n)$  significa que é da mesma ordem de convergência de  $1/n$ . Nesta situação, diz-se que o estimador é assintoticamente equivalente ao funcional, o que permite que, na prática, se ultrapasse a questão teórica.

*Exemplo 4.1.2.* O valor médio de uma população, com função distribuição de probabilidade  $F$ , é definido pelo funcional

$$\mu(F) = T(F) = \int x dF(x).$$

Deste modo,  $\mu(F) = \int x f_\theta(x) dx$ , se a distribuição admitir função densidade de probabilidade  $f_\theta$  e  $\mu(F) = \sum_{i=1}^n x_i f_\theta(x_i)$ , se a distribuição admitir função probabilidade  $f_\theta$ . Note-se que  $F$  é um dos elementos do espaço de funções  $\mathcal{F}$ . Para outro elemento de  $\mathcal{F}$ , por exemplo, uma outra função distribuição de probabilidade  $G$ , o funcional  $T$  permite obter a média da distribuição  $G$ , calculando  $T$  na distribuição  $G$ , ou seja,  $T(G) = \mu(G)$ . Do mesmo modo, para a função distribuição empírica, tem-se que  $G_n \in \mathcal{F}$  e calculando  $T(G_n) = \mu(G_n)$ , obtém-se a média dessa distribuição que, neste caso, é a média amostral:

$$\begin{aligned} \mu(G_n) &= T(G_n) = \int x dG_n(x) \\ &= \sum_{i=1}^n x_i \times \frac{1}{n} = \bar{x}, \end{aligned}$$

pois  $G_n$  é uma distribuição discreta, em que todos os valores da amostra,  $x_i$ , são equiprováveis. Assim, o mesmo funcional pode ser calculado em  $F$ , resultando na média da população, ou em  $G_n$ , resultando na média da amostra.  $\diamond$

Para ilustrar uma situação em que um estimador apenas é assintoticamente equivalente a um funcional estatístico, considera-se no exemplo seguinte o caso da variância amostral corrigida.

*Exemplo 4.1.3.* A variância de uma população, com função distribuição de probabilidade  $F$ , é definida pelo funcional  $\sigma^2$ , do seguinte modo

$$\sigma^2(F) = T(F) = \int [x - \mu(F)]^2 dF(x),$$

onde  $\mu(F)$  foi definido no exemplo anterior. Calculando o funcional  $\sigma^2$  na distribuição empírica  $G_n$ , vem

$$\begin{aligned} \sigma^2(G_n) &= \int [x - \mu(G_n)]^2 dG_n(x) \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 \times \frac{1}{n} \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2. \end{aligned}$$

Denote-se por  $S_c^2$  a variância amostral corrigida, com  $S_c^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ . Tendo em conta que  $S_c^2 = \frac{n}{n-1} \sigma^2(G_n)$ , note-se que  $S_c^2$  não depende de  $n$  apenas através de  $G_n$ , logo não admite uma representação como um funcional da distribuição empírica. No entanto, como a diferença  $S_c^2 - \sigma^2(G_n) = \frac{1}{n-1} \sigma^2(G_n)$ , vem que

$$\lim_{n \rightarrow \infty} \frac{\frac{\sigma^2(G_n)}{n-1}}{\frac{1}{n}} = \lim_{n \rightarrow \infty} \frac{n}{n-1} \sigma^2(G_n) = \sigma^2(G_n),$$

donde se conclui que  $S_c^2$  é assintoticamente equivalente ao funcional  $\sigma^2$ .  $\diamond$

Quando o estimador é interpretado como um funcional estatístico há necessidade de adaptar a noção de consistência. A noção que lhe corresponde é a de consistência à Fisher do respectivo funcional. Um funcional  $T$ , diz-se consistente à Fisher, no modelo  $F_\theta$ , quando

$$T(F_\theta) = \theta, \forall \theta \in \Theta.$$

Sob certas condições de regularidade, a consistência à Fisher é equivalente à consistência usual (ver Pires e Branco, 2007).

A partir da representação de um estimador como um funcional, é possível apresentar uns dos principais instrumentos de avaliação da robustez - a função de influência - que se passa a definir.

**Definição 4.1.2.** A **função de influência** ( $IF$ ) do funcional  $T$  na distribuição  $F$  é definida por

$$IF(x, T, F) = \lim_{\delta \rightarrow 0} \frac{T((1 - \delta)F + \delta\Delta_x) - T(F)}{\delta},$$

para todos os pontos  $x$  em que o limite existe, onde  $\delta > 0$ ,  $F \in \mathcal{D}(T)$  e  $\Delta_x$  representa a função distribuição discreta degenerada em  $x$ , i.e.,  $\Delta_x(u) = 1$ , se  $u \geq x$  e  $\Delta_x(u) = 0$ , se  $u < x$ .  $\square$

Note-se que  $(1 - \delta)F + \delta\Delta_x$  representa a distribuição de uma variável aleatória que é uma mistura, nas proporções de  $1 - \delta$  para  $\delta$ , de duas variáveis aleatórias com distribuição  $F$  e  $\Delta_x$ , respectivamente. Ou seja, considera-se um modelo de contaminação em  $x$ , em que a probabilidade de encontrar uma observação proveniente do verdadeiro modelo  $F$ , é  $1 - \delta$  e a probabilidade de se observar  $x$  é  $\delta$ . Assim, a função de influência permite avaliar o efeito que a inclusão de uma observação  $x$  (com pequena contaminação  $\delta$ ) provoca sobre as estimativas, quando as restantes observações provêm da distribuição suposta verdadeira.

A ideia de base da função de influência é idêntica à da curva de sensibilidade, mas agora num contexto assintótico, ou seja, para grandes amostras. Pode mostrar-se que  $SC_n(x, T) \simeq IF(x, T, F)$  e que  $SC_n(x, T) \xrightarrow[n \rightarrow \infty]{p} IF(x, T, F)$ , desde que o estimador  $T$  seja pelo menos assintoticamente equivalente a um funcional regular  $T$ .

Atendendo à expressão da  $IF$ , esta pode ser vista como a derivada do funcional que define o estimador, ou seja,

$$IF(x, T, F) = \left. \frac{\partial [T(F_{\delta,x})]}{\partial \delta} \right|_{\delta=0},$$

com  $F_{\delta,x} = (1 - \delta)F + \delta\Delta_x$ . Esta forma de apresentar a  $IF$  pode ser mais conveniente em termos de cálculo e pode ser interpretada como sendo a taxa de variação do estimador perante uma quantidade  $\delta$  de contaminação.

Referem-se a seguir algumas propriedades da função de influência.



PIF.1 Na distribuição  $F$ , a função de influência tem esperança nula, isto é

$$E_F [IF(X, T, F)] = \int IF(x, T, F) dF(x) = 0.$$

PIF.2 Sob condições de regularidade, a esperança  $E_F [IF^2(X, T, F)] = V(T, F)$ , onde  $V(T, F)$  é a variância assintótica do estimador  $T$ .

PIF.3 Para uma função distribuição  $G$  próxima de  $F$  tem-se que

$$T(G) \simeq T(F) + \int IF(x, T, F) dG(x).$$

Esta expressão é obtida com os termos de primeira ordem resultantes do desenvolvimento em série de Taylor.

PIF.4 Seja  $(X_1, \dots, X_n)$  uma amostra aleatória duma população com distribuição  $F$ . Sob condições de regularidade, o Teorema do Limite Central garante que

$$\sqrt{n}(T_n - T(F)) \overset{\bullet}{\sim} N(0, V(T, F)).$$

Este resultado mostra que, em muitos casos, a  $IF$  permite também deduzir a distribuição assintótica do estimador.

Para além das propriedades já referidas, podem ainda definir-se várias medidas de robustez com base na função de influência.

**Definição 4.1.3.** Define-se a **sensibilidade a grandes erros** de  $T$  em  $F$  por

$$\gamma^* = \gamma^*(T, F) = \sup_x |IF(x, T, F)|.$$

□

É desejável que  $\gamma^*$  assuma um valor baixo.

**Definição 4.1.4.**  $T$  é **B-robusto** em  $F$  se  $\gamma^*(T, F) < \infty$ .

□

Pode dizer-se que um estimador é B-robusto se tiver função de influência limitada.

Há ainda outras medidas de robustez baseadas na função de influência que, embora sejam menos utilizadas, não queremos deixar de apresentar numa introdução ao tema.

**Definição 4.1.5.** Chama-se **sensibilidade local** de  $T$  em  $F$  ao valor

$$\lambda^*(T, F) = \sup_{x \neq y} \frac{|IF(y, T, F) - IF(x, T, F)|}{|y - x|}.$$

□

Este valor mede o pior efeito possível (estandardizado) que variações locais nas observações podem provocar na estimativa. Este tipo de variações pode ocorrer quando se procede ao arredondamento ou ao agrupamento de dados; é conveniente que esta medida tenha um valor reduzido.

**Definição 4.1.6.** Para uma distribuição  $F$  simétrica em relação a um ponto  $c$ , define-se **ponto de rejeição** de  $T$  em  $F$  como sendo

$$\rho^*(T, F) = \inf \{r > 0 : IF(x, T, F) = 0 \text{ para } x \text{ tal que } |x - c| > r\}.$$

□

O ponto de rejeição indica a menor distância ao centro de simetria a partir da qual a função de influência é identicamente nula, isto é, indica a distância a partir da qual devem desprezar-se as observações, passando a não interferir no valor da estimativa. É desejável que este valor seja finito.

Resumindo, para que um estimador tenha boas propriedades de robustez, com base na sua função de influência, deve verificar os seguintes requisitos:

- Ter função de influência limitada, isto é,  $\gamma^*(T, F_\theta) < \infty$ ,
- $E_F[IF^2(X, T, F)] \simeq \mathcal{I}(\theta)^{-1}$  (onde  $\mathcal{I}(\theta)$  é a quantidade de informação de Fisher) para que a eficiência assintótica seja elevada,
- $\lambda^*(T, F_\theta) < \infty$ ,
- $\rho^*(T, F_\theta) < \infty$ .

A abordagem da robustez baseada no estudo da função de influência é designada por abordagem infinitesimal, ou designada por conceito infinitesimal da robustez. Há outros

conceitos de robustez na literatura, nomeadamente, a abordagem qualitativa, que se baseia na continuidade do funcional, e a abordagem quantitativa, que se baseia na noção de ponto de ruptura. A abordagem qualitativa não tem tido grande desenvolvimento, eventualmente pelo difícil tratamento matemático que exige. Em todo o caso, faz-se uma breve apresentação do conceito. A abordagem quantitativa é apresentada no final da presente secção e tem vindo a motivar cada vez mais investigadores. Como é de esperar, os três conceitos de robustez estão interligados e os respectivos indicadores estão relacionados entre si.

A robustez qualitativa está associada à continuidade do estimador, no sentido em que um estimador é robusto se dadas duas distribuições próximas,  $F$  e  $G$ , as estimativas baseadas em amostras obtidas dessas populações também estiverem muito próximas.

**Definição 4.1.7.** Uma sucessão de estimadores  $(T_n)_{n \in \mathbb{N}}$  é **qualitativamente robusta** se

$$\forall \varepsilon > 0, \exists \delta > 0, n_0 : \forall F \in \mathcal{F}, \forall n > n_0, d(F, G) < \delta \Rightarrow d(\mathcal{L}_F(T_n), \mathcal{L}_G(T_n)) < \varepsilon,$$

onde  $d(F, G)$  representa uma distância adequada no espaço das distribuições e  $\mathcal{L}_F(T_n)$  representa a distribuição de  $T_n$  quando as observações têm distribuição  $F$ .  $\square$

Este conceito encontra-se intimamente relacionado com o de continuidade de um estimador.

**Definição 4.1.8.** Uma **sucessão** de estimadores  $(T_n)_{n \in \mathbb{N}}$  é **contínua** em  $F$  se

$$\begin{aligned} \forall \varepsilon > 0, \exists \delta > 0, n_0 : \forall m, n > n_0, \forall F_n F_m, \\ [d(F, F_n) < \delta \wedge d(F, F_m) < \delta] \Rightarrow |T_n(F_n) - T_m(F_m)| < \varepsilon. \end{aligned}$$

$\square$

Tendo em conta que os critérios de robustez relacionados com a função de influência são estritamente locais, na medida em que descrevem o comportamento do estimador quando se introduz uma pequena percentagem de contaminação, interessa complementar o estudo de robustez com uma medida global, que indique qual a maior percentagem

de contaminação que o estimador admite antes de começar a produzir estimativas disparatadas. Chama-se a este valor o ponto de ruptura, o qual constitui o principal indicador da robustez, de acordo com a abordagem quantitativa.

A noção de ponto de ruptura é muito intuitiva na sua versão amostral, ou seja, em dimensão finita. É deste modo que se avalia a robustez pela abordagem quantitativa na maior parte dos estudos. Tomando o limite em  $n$  no ponto de ruptura empírico, obtém-se o ponto de ruptura do estimador, cuja definição matemática e respectiva interpretação são bem mais delicadas.

**Definição 4.1.9.** A partir da realização amostral  $(x_1, \dots, x_n)$  substituam-se quaisquer  $m$  observações,  $(x_{i_1}, \dots, x_{i_m})$ , por  $m$  valores arbitrários  $(y_1, \dots, y_m)$  e designe-se a amostra resultante por  $(z_1, \dots, z_n)$ . O **ponto de ruptura empírico (ou em dimensão finita)** de  $T$  é dado por

$$\varepsilon_n^*(T) = \varepsilon_n^*(T, x_1, \dots, x_n) = \frac{1}{n} \max \left\{ m : \max_{i_1, \dots, i_m} \sup_{y_1, \dots, y_m} |T(z_1, \dots, z_n)| < \infty \right\}.$$

□

Para apresentar a definição de ponto de ruptura (não empírico), é necessário introduzir previamente a noção de enviesamento assintótico máximo.

**Definição 4.1.10.** Considere-se uma vizinhança de contaminação  $\delta$  da função distribuição  $F$ , definida por Huber, como

$$\mathcal{P}_\delta(F) = \{G : G = (1 - \delta)F + \delta H, H \in \mathcal{F}\}.$$

Chama-se **enviesamento assintótico máximo** de  $T$  na vizinhança  $\mathcal{P}_\delta(F)$  a

$$b_{T,F}(\delta) = \sup_{G \in \mathcal{P}_\delta(F)} |T(G) - T(F)|.$$

□

Este valor corresponde ao erro máximo que se pode cometer ao estimar  $T(F)$ , a partir de uma amostra contaminada por outra distribuição  $H$ , com uma percentagem de contaminação  $\delta$ . Interessa que este valor seja o mais reduzido possível.

**Definição 4.1.11.** O ponto de ruptura assintótico do estimador  $T$  é definido por

$$\varepsilon^*(T, F) = \sup \{ \delta : b_{T,F}(\delta) < b_{T,F}(1) \}.$$

□

Numa perspectiva de robustez, é conveniente que um estimador tenha ponto de ruptura estritamente positivo e deve ser o mais elevado possível. Note-se, no entanto, que o maior ponto de ruptura com sentido prático é de 50%. Logo, interessa que um estimador tenha um ponto de ruptura próximo de 50%.

Tendo em conta que foram definidos vários indicadores que permitem avaliar a robustez de um estimador, interessa ter uma ideia da ordem pela qual devem ser utilizados. De acordo com as tendências mais recentes (Pires e Branco, 2007), considera-se prioritária a avaliação da robustez pelo ponto de ruptura, seguindo-se o estudo da eficiência assintótica e do enviesamento assintótico máximo.

## 4.2 Estimação robusta

Esta secção dedica-se principalmente ao estudo dos estimadores-M. Estes estimadores foram propostos por Huber (1964), como uma forma de generalizar os estimadores de máxima verosimilhança, o que justifica o seu nome, e tiveram um papel importante no desenvolvimento da robustez. São ainda apresentadas outras classes de estimadores. A secção começa por relembrar princípios de invariância e de equivariância, que devem estar presentes no processo de construção de estimadores, em particular dos estimadores-M.

### 4.2.1 Invariância e equivariância

Em qualquer aplicação do princípio da invariância, uma função  $T(x)$  é especificada, mas, se  $T(x) = T(y)$ , então o princípio da invariância afirma que a inferência feita se  $x$  é observado deve ter uma certa relação com a inferência feita se  $y$  é observado, embora as duas inferências possam não ser iguais. Esta restrição sobre o procedimento de inferência, pode conduzir a uma simplificação da análise de dados. O princípio de

invariância combina dois tipos diferentes de invariância. O primeiro tipo de invariância, que pode ser chamado de invariância da medição, estabelece que a inferência feita não deve depender da escala que é utilizada na medição. O segundo tipo de invariância, que se pode chamar invariância formal, afirma que se dois problemas de inferência têm a mesma estrutura formal, em termos do modelo matemático utilizado, então deve ser utilizado o mesmo procedimento de inferência em ambos os problemas (mais detalhes podem ser vistos em Casella e Berger, 1990). A aplicação do princípio da invariância a diferentes problemas de inferência pode conduzir a efeitos diferentes. Para observar essas aplicações, interessa definir alguns tipos de modelos.

**Definição 4.2.1.** Um **modelo de localização**, com parâmetro de localização  $\theta$ , consiste numa família de distribuições  $\{F_\theta, \theta \in \mathbb{R}\}$  tal que

$$F_\theta(x) = F(x - \theta),$$

onde  $F = F_0$  representa a distribuição univariada genérica que define a família do modelo.  $\square$

Sendo  $f_\theta$  a função probabilidade ou a função densidade associada a um modelo de localização  $F_\theta$  (consoante o caso), então verifica-se que  $f_\theta(x) = f_\theta(x - \theta)$ .

**Definição 4.2.2.** Um **modelo de escala**, com parâmetro de escala  $\theta$ , é uma família de distribuições  $\{F_\theta, \theta \in \mathbb{R}^+\}$  tal que

$$F_\theta(x) = F(x/\theta),$$

onde  $F = F_1$  corresponde à distribuição univariada genérica que define a família do modelo.  $\square$

Note-se que quando existe função densidade,  $f = F'$ ,  $f(x)$  representa a densidade para  $\theta = 1$ . A função densidade para um valor genérico  $\theta$ , verifica

$$f_\theta(x) = \frac{\partial}{\partial \theta} [F(x/\theta)] = (1/\theta) f(x/\theta).$$

Ao estimar uma medida de localização, é importante usar um estimador que, quando calculado para uma amostra resultante da translacção dos dados, apresente um valor

que se relacione através da mesma transformação com o valor que o estimador assume para o conjunto original de dados. Esta propriedade é chamada equivariância de localização.

Quando se estima uma medida de dispersão, é importante usar um estimador tal que, quando se multiplicam todas as observações da amostra por uma constante positiva, o valor do estimador sofra a mesma alteração. Esta propriedade é chamada equivariância de escala. Um estimador de escala com esta propriedade produz um resultado equivalente, independentemente da unidade de medida em que se encontrem as observações.

Apresentam-se de seguida as definições de equivariância para os diferentes tipos de modelos.

**Definição 4.2.3.** Um estimador  $T_n$  de um parâmetro de localização é **equivariante em relação à localização** se, dada uma amostra aleatória  $(X_1, \dots, X_n)$ ,

$$T_n(X_1 + k, \dots, X_n + k) = T_n(X_1, \dots, X_n) + k, \forall k \in \mathbb{R} \quad (4.2.1)$$

e é **equivariante em relação à escala** se verificar

$$T_n(kX_1, \dots, kX_n) = kT_n(X_1, \dots, X_n), \forall k \in \mathbb{R} \setminus \{0\}.$$

□

**Definição 4.2.4.** Um estimador  $Q_n$  de um parâmetro de escala é **equivariante em relação à localização** se, dada uma amostra aleatória  $(X_1, \dots, X_n)$ ,

$$Q_n(X_1 + k, \dots, X_n + k) = Q_n(X_1, \dots, X_n), \forall k \in \mathbb{R} \quad (4.2.2)$$

e é **equivariante em relação à escala** se verificar

$$Q_n(kX_1, \dots, kX_n) = kQ_n(X_1, \dots, X_n), \forall k \in \mathbb{R}^+. \quad (4.2.3)$$

□

Pode designar-se o efeito descrito em (4.2.1) como **equivariância**, na medida em que o estimador sofre alterações de acordo com a forma como os dados são transformados. Embora alguns autores designem esta propriedade por invariância, outros

reservam o termo **invariância** para situações do tipo da descrita em (4.2.2), na qual o estimador se mantém inalterado quando os dados são transformados.

Como exemplo do caso multiparamétrico, refere-se a seguir o modelo de localização e escala, generalizando os modelos de localização e de escala apresentados anteriormente.

Para juntar num só conceito os modelos de localização e de escala, pode partir-se de uma variável aleatória  $X$  e considerar uma transformação linear que corresponda a uma translação e conjuntamente a uma mudança de escala,  $X_{\mu,\sigma} = \sigma X + \mu$ . A função de distribuição relativa a esta nova variável é  $F_{X_{\mu,\sigma}}(x) = P(\sigma X + \mu \leq x) = P\left(X \leq \frac{x - \mu}{\sigma}\right) = F\left(\frac{x - \mu}{\sigma}\right)$ .

Como exemplo, refira-se o caso da distribuição normal, uma vez que se  $X \sim N(\mu, \sigma^2)$ , então  $\frac{X - \mu}{\sigma} \sim N(0, 1)$ , isto é, ainda tem distribuição normal.

**Definição 4.2.5.** Um **modelo de localização e escala**, com parâmetro de localização  $\mu$  e parâmetro de escala  $\sigma$ , consiste numa família de distribuições  $\{F_\theta, \theta \in \Theta\}$ , onde  $\theta = \begin{bmatrix} \mu \\ \sigma \end{bmatrix}$  e  $\Theta = \mathbb{R} \times \mathbb{R}^+$ , tal que

$$F_\theta(x) = F\left(\frac{x - \mu}{\sigma}\right),$$

onde  $F = F_{0,1}$  corresponde à distribuição univariada genérica que define a família do modelo.  $\square$

Importa salientar que os parâmetros de localização e de escala,  $\mu$  e  $\sigma$ , podem não coincidir com o valor médio e o desvio padrão.

Estas noções de invariância podem ser adaptadas aos modelos de regressão, como se verá de seguida.

**Definição 4.2.6.** Seja um modelo de regressão linear múltipla definido por

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

onde  $\mathbf{X}$  é uma matriz  $(n \times p)$  das observações dos regressores,  $\mathbf{y} \in \mathbb{R}^n$  é um vector de observações de uma variável resposta,  $\boldsymbol{\beta} \in \mathbb{R}^p$  é um vector de parâmetros desconhecidos e  $\boldsymbol{\varepsilon} \in \mathbb{R}^n$  é o vector dos erros do modelos. Um estimador  $\hat{\boldsymbol{\beta}}$  do parâmetro  $\boldsymbol{\beta}$  é



**equivariante em relação à regressão** se verificar

$$\hat{\beta}(\mathbf{X}, k\mathbf{y}) = k\hat{\beta}(\mathbf{X}, \mathbf{y}), \forall k \in \mathbb{R};$$

**é equivariante em relação à escala** se verificar

$$\hat{\beta}(\mathbf{X}, \mathbf{y} + \mathbf{X}\boldsymbol{\lambda}) = \hat{\beta}(\mathbf{X}, \mathbf{y}) + \boldsymbol{\lambda}, \forall \boldsymbol{\lambda} \in \mathbb{R}^p;$$

e é **equivariante afim** se verificar

$$\hat{\beta}(\mathbf{A}\mathbf{X}, \mathbf{y}) = \mathbf{A}^{-1}\hat{\beta}(\mathbf{X}, \mathbf{y}),$$

para qualquer matriz  $\mathbf{A}_{(p \times p)}$  não singular.  $\square$

### 4.2.2 Estimadores-M

Num modelo paramétrico com função densidade (ou função de probabilidade)  $f(x, \theta)$ , o estimador de máxima verosimilhança (*EMV*) para  $\theta$ , obtido para uma amostra aleatória, é o estimador  $T_n = T_n(X_1, \dots, X_n)$  que maximiza a função de verosimilhança  $\prod_{i=1}^n f(X_i, T_n)$ . Sob condições de regularidade, em geral, o *EMV* é determinado maximizando  $\sum_{i=1}^n \log f(X_i, T_n)$ , ou ainda, minimizando  $\sum_{i=1}^n [-\log f(X_i, T_n)]$ .

Para obter  $T_n$  resolve-se a equação resultante da condição de primeira ordem,

$$\begin{aligned} \sum_{i=1}^n -\frac{\partial \log f(X_i, T_n)}{\partial \theta} &= 0 \\ \Leftrightarrow \sum_{i=1}^n \psi(X_i, T_n) &= 0, \end{aligned} \quad (4.2.4)$$

com  $\psi(x, \theta) = -\frac{\partial \log f(x, \theta)}{\partial \theta}$  e confirma-se que a solução é um mínimo recorrendo às condições de segunda ordem.

O estimador-M vai generalizar o estimador de máxima verosimilhança, seguindo o mesmo princípio, mas podendo usar outras funções  $\psi$ , definidas de forma a construir estimadores robustos.

**Definição 4.2.7.** Um **estimador-M** é um estimador  $T_n$  determinado pela solução do problema

$$\min \sum_{i=1}^n \rho(X_i, T_n), \quad (4.2.5)$$

onde  $\rho$  é a função objectivo, a minimizar. Se existir  $\psi(x, \theta) = \frac{\partial \rho(x, \theta)}{\partial \theta}$ , então  $T_n$  é solução da equação (4.2.4).  $\square$

Neste sentido, o *EMV* é um estimador-M para o caso particular de  $\psi(x, \theta) = -\frac{\partial \log f(x, \theta)}{\partial \theta}$  e  $\rho(x, \theta) = -\log f(x, \theta)$ . Note-se que, de acordo com a **Definição 4.2.7**, um estimador-M não é sempre um estimador robusto, tudo vai depender das funções  $\psi$  e  $\rho$ .

O processo mais usado para obter um estimador-M é através da resolução de (4.2.4), embora só seja equivalente a resolver (4.2.5), sob certas condições de regularidade. No entanto, isso acontece para a maior parte das distribuições usadas nas aplicações. Por outro lado, a expressão (4.2.4) facilita a generalização ao caso multivariado. Sendo assim, é usual identificar-se o estimador-M com a função  $\psi$  que o define.

O funcional  $T$  equivalente a  $T_n$  expresso como um estimador-M, vem definido implicitamente na distribuição  $F$  por

$$\int \psi(x, T(F)) dF(x) = 0. \quad (4.2.6)$$

Um dos principais motivos que levou à utilização dos estimadores-M nas investigações sobre robustez é o facto da sua função de influência ser proporcional à função  $\psi$  que o define. Na realidade, prova-se que

$$IF(x, T, F) = \frac{\psi(x, T(F))}{-\int \left[ \frac{\partial}{\partial \theta} \psi(u, \theta) \right]_{\theta=T(F)} dF(u)}. \quad (4.2.7)$$

Esta propriedade permite que se definam estimadores robustos por construção, considerando funções  $\psi$  limitadas, pois a função de influência do estimador, sendo proporcional a  $\psi$  também será limitada e, conseqüentemente, o respectivo estimador é robusto.

De seguida consideram-se separadamente os estimadores-M em modelos de localização e de escala.

### Modelo de localização

Dada uma amostra aleatória  $(X_1, \dots, X_n)$ , o *EMV* para o parâmetro  $\theta$  do modelo de localização é a solução do problema

$$\min \sum_{i=1}^n \rho(X_i - T_n),$$

com  $\rho(u) = -\log f(u)$ , o que equivale, em situações de regularidade, a ser solução do problema

$$\sum_{i=1}^n \psi(X_i - T_n) = 0, \quad (4.2.8)$$

onde  $\psi(u) = \frac{\partial \rho(u)}{\partial u} = -\frac{f'(u)}{f(u)}$ .

Considerando um modelo de localização, sabe-se que  $\psi(x, \theta) = \psi(x - \theta)$  e define-se o estimador-M para um modelo de localização através da condição (4.2.8).

Para que o funcional  $T$ , definido por  $\int \psi(x - T(F)) dF(x) = 0$ , seja consistente à Fisher, deve ter-se

$$\int \psi(x - \theta) dF(x - \theta) = \int \psi(u) dF(u) = 0.$$

Esta condição verifica-se sempre que a distribuição  $F$  seja simétrica em relação à origem e  $\psi$  seja um função ímpar.

Pode verificar-se que um estimador-M definido por (4.2.8) é equivariante em relação à localização, ou seja, verifica a condição (4.2.1), e que

$$IF(x, T, F_\theta) = IF(x - \theta, T, F) = IF(u, T, F),$$

com  $u = x - \theta$ . Este resultado traduz que todas as propriedades do estimador que se baseiam na função de influência podem ser estudadas relativamente à distribuição central  $F$ .

Como se afirma por exemplo em Pires e Branco (2007) ou em Hampel *et al.* (1986), no caso de  $\psi$  ser não decrescente, se  $\psi$  for limitada, o estimador-M é qualitativamente robusto e tem ponto de ruptura máximo  $\varepsilon^*(T, F) = 1/2$ ; se  $\psi$  for ilimitada, o estimador não é qualitativamente robusto e tem ponto de ruptura  $\varepsilon^*(T, F) = 0$ . Por considerações deste tipo se vê a importância de seleccionar funções  $\psi$  (ou funções  $\rho$ ) com propriedades convenientes para construir um estimador robusto.

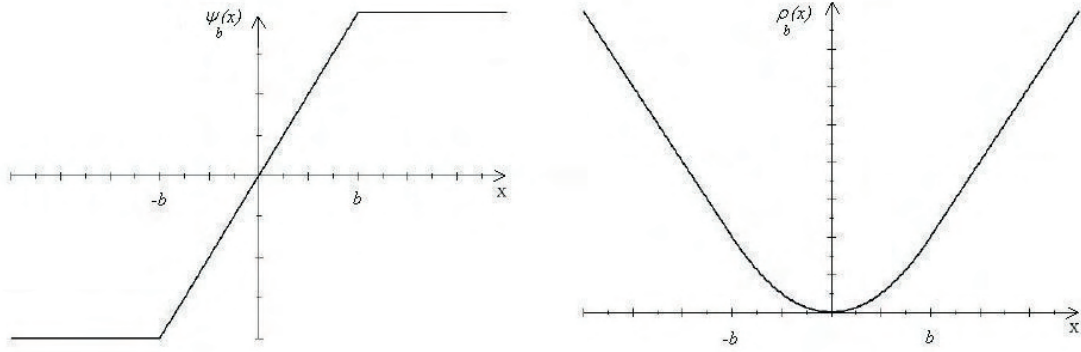


Figura 4.1: Gráficos das funções  $\Psi$  e  $\rho$  de Huber associadas ao estimador-M de localização.

Apresentam-se de seguida alguns exemplos de funções  $\psi$  e  $\rho$ , para estimar parâmetros de localização em distribuições normais.

- (i) As funções  $\psi$  e  $\rho$  de Huber associadas ao estimador-M de Huber são, respectivamente

$$\psi_b(x) = \begin{cases} -b & , x < -b \\ x & , |x| \leq b \\ b & , x > b \end{cases} ; \rho_b(x) = \begin{cases} \frac{x^2}{2} & , |x| \leq b \\ b|x| - \frac{b^2}{2} & , |x| > b \end{cases} . \quad (4.2.9)$$

e os respectivos gráficos encontram-se na Figura 4.1.

- (ii) As funções  $\psi$  e  $\rho$  bponderada de Tukey associadas ao estimador-M de Tukey são, respectivamente

$$\psi(x) = \begin{cases} x(r^2 - x^2)^2 & , |x| \leq r \\ 0 & , |x| > r \end{cases} , \rho(x) = \begin{cases} \frac{x^6}{6} - \frac{r^2 x^4}{2} + \frac{r^4 x^2}{2} & , |x| \leq r \\ \frac{r^6}{6} & , |x| > r \end{cases} .$$

Huber propôs as primeiras funções  $\psi$  e  $\rho$ , definidas em (i), quando apresentou a definição de estimador-M; posteriormente outras funções  $\psi$  foram propostas por outros autores e podem ser encaradas como alterações feitas na função proposta por Huber, de modo a conduzir a estimadores com melhores propriedades.

Note-se que a utilização de um estimador-M exige que se escolha a função  $\psi$  e eventualmente os valores das constantes que possam entrar na sua definição, que se designam por constantes de afinação. Para cada função, a escolha das constantes vai regular a eficiência e a robustez do estimador, havendo necessidade de equilibrar eficiência e robustez para cada modelo.

De um modo geral, os estimadores-M não são equivariantes em relação à escala. Para ultrapassar esta dificuldade usa-se um estimador auxiliar de escala que seja equivariante em relação à escala. Deste modo, passaram a utilizar-se estimadores-M, definidos como soluções de

$$\sum_{i=1}^n \psi \left( \frac{X_i - T_n}{S_n} \right) = 0. \quad (4.2.10)$$

É frequente usar-se o *MAD* como estimador auxiliar  $S_n$ , pelo facto de ser robusto. A questão da baixa eficiência do *MAD* sob o modelo normal não é relevante, pelo facto do objectivo ser a estimação do parâmetro de localização. Para além disso, esse aspecto pode ser melhorado pela multiplicação de uma constante, tendo em conta que se  $T_n$  é o estimador obtido como solução de (4.2.10), então  $kT_n$  é solução da equação obtida substituindo  $X_i$  por  $kX_i$  em (4.2.10), com  $k > 0$ .

Na maior parte dos casos, o cálculo das estimativas é feito recorrendo a métodos iterativos. Assim, é importante que se use um estimador altamente robusto, como por exemplo a mediana, para a escolha de um valor inicial, de modo a assegurar uma rápida convergência para a solução. Para certas funções  $\rho$ , o problema de minimização tem várias soluções, correspondentes a vários zeros de (4.2.10), sendo, nesses casos, ainda mais importante a escolha de uma boa solução inicial.

Outro modo de obter as estimativas-M consiste em calcular, em cada iteração, uma média ponderada das observações, com pesos que variam de iteração para iteração, de acordo com

$$T_n^{(k+1)} = \frac{\sum_{i=1}^n x_i w \left( u_i^{(k)} \right)}{\sum_{i=1}^n w \left( u_i^{(k)} \right)}, \quad (4.2.11)$$

onde  $u_i^{(k)} = \frac{x_i - T_n^{(k)}}{S_n}$ .

Como se refere em Pires e Branco (2007), se

$$w(u) = \begin{cases} \frac{\psi(u)}{u} & , u \neq 0 \\ 1 & , u = 0 \end{cases}$$

e a sucessão das estimativas for convergente, então o limite é a solução de (4.2.10).

Efectivamente, ao substituir  $\psi(u) = uw(u)$  em (4.2.10), obtém-se

$$\begin{aligned} \sum_{i=1}^n w\left(\frac{X_i - T_n}{S_n}\right) \times \left(\frac{X_i - T_n}{S_n}\right) &= 0 \\ \Leftrightarrow T_n &= \frac{\sum_{i=1}^n X_i w\left(\frac{X_i - T_n}{S_n}\right)}{\sum_{i=1}^n w\left(\frac{X_i - T_n}{S_n}\right)}, \end{aligned}$$

donde resulta que uma forma alternativa para construir estimadores-M é através da definição da função  $w(u)$ .

Os estimadores obtidos de acordo com (4.2.11) podem ser interpretados como um generalização dos estimadores dos mínimos quadrados, são muito utilizados em regressão e normalmente são conhecidos por mínimos quadrados iterativamente ponderados ou *IRWLS* (de *Iteratively Reweighted Least Squares*). Alguns autores, por exemplo, Rousseeuw e Croux (1994), defendem que se obtêm boas estimativas apenas com um ou dois passos daqueles algoritmos. Estimadores definidos iterativamente por ponderações são também designados genericamente por **estimadores-W**.

Na prática, a obtenção das estimativas pode ser feita com recurso ao programa *R*, em particular usando as *packages robustbase* e *MASS*.

Por fim, importa saber de entre todos os estimadores-M qual se deve escolher para cada situação. A função  $\psi$  de Huber foi a mais utilizada durante muito tempo mas, actualmente, parece haver maior tendência para funções  $\psi$  com ponto de rejeição finito, ou decrescente a partir de certo ponto. Assim, a função  $\psi$  de Tukey tem vindo a ser mais usada exigindo, no entanto, um grande cuidado na escolha da solução inicial.

## Modelo de escala

Dada uma amostra aleatória  $(X_1, \dots, X_n)$ , o *EMV*,  $S_n = S_n(X_1, \dots, X_n)$ , do parâmetro  $\theta$  do modelo de escala é a solução do problema

$$\min \left[ n \log S_n + \sum_{i=1}^n \rho\left(\frac{X_i}{S_n}\right) \right],$$

com  $\rho(u) = -\log f(u)$ . Sob condições regularidade, resolver este problema é equivalente à resolução do seguinte

$$\sum_{i=1}^n \psi\left(\frac{X_i}{S_n}\right) = 0, \quad (4.2.12)$$

onde  $\psi(u) = \frac{\partial \rho(u)}{\partial u} = -u \frac{f'(u)}{f(u)} - 1$ . Pode reescrever-se (4.2.12) na forma

$$\sum_{i=1}^n \chi\left(\frac{X_i}{S_n}\right) = n, \quad (4.2.13)$$

onde  $\chi(u) = -u \frac{f'(u)}{f(u)} = \psi(u) + 1 \Leftrightarrow \psi(u) = \chi(u) - 1$ .

Define-se um estimador-M para um modelo de escala, por generalização do *EMV*, através da condição (4.2.13), considerando que  $\psi(x, \theta) = \psi\left(\frac{x}{\theta}\right)$ .

Para que o funcional  $S$ , definido por  $\int \psi\left(\frac{x}{S(F)}\right) dF(x) = 0$ , seja consistente segundo Fisher, deve ter-se que

$$\begin{aligned} \int \psi\left(\frac{x}{\theta}\right) dF\left(\frac{x}{\theta}\right) &= 0 \\ \Leftrightarrow \int \psi(u) dF(u) &= 0 \\ \Leftrightarrow \int (\chi(u) - 1) dF(u) &= 0. \end{aligned}$$

Nem sempre esta condição origina um estimador consistente, como acontece, por exemplo, no caso de  $\chi$  não ser limitada. Assim, para garantir a consistência do estimador  $S$ , uma vez fixada a forma de  $\chi$ , define-se

$$\int \chi(u) dF(u) = b \quad (4.2.14)$$

e considera-se  $\psi(u) = \chi(u) - b$  na definição do estimador.

Deste modo, dada uma amostra aleatória  $(X_1, \dots, X_n)$ , um estimador-M,  $S_n = S_n(X_1, \dots, X_n)$ , do parâmetro de escala do modelo de escala, é solução de

$$\sum_{i=1}^n \chi\left(\frac{X_i}{S_n}\right) = nb, \quad (4.2.15)$$

onde  $\chi$  é uma função genérica e  $b$  é uma constante definida por (4.2.14).

Uma propriedade desejável para os estimadores do parâmetro de escala de um modelo de escala é a equivariância em relação à escala. Os estimadores definidos por

(4.2.15) gozam desta propriedade, ou seja, verificam a condição (4.2.3), e que

$$IF(x, S, F_\theta) = \theta IF\left(\frac{x}{\theta}, S, F\right) = \theta IF(u, S, F),$$

com  $u = \frac{x}{\theta}$ . A partir deste último resultado, todas as propriedades do estimador, que se estabelecem a partir da função de influência, podem ser baseadas na distribuição central  $F$ .

Se  $\psi$  for não decrescente para  $x > 0$ , se  $\psi$  for limitada pode mostrar-se (ver Hampel *et al.*, 1986) que o estimador  $S$  do parâmetro de escala é qualitativamente robusto e tem ponto de ruptura máximo  $\varepsilon^*(S, F) = \min \left\{ \frac{1}{2}, \frac{b}{\chi(\infty)} \right\}$ ; no caso de  $\psi$  ser ilimitada, o estimador não é qualitativamente robusto e tem ponto de ruptura  $\varepsilon^*(S, F) = 0$ .

Quando se admite que as observações seguem aproximadamente um modelo de escala normal, sabe-se que a função  $\chi$  deve verificar um conjunto de características específicas, podendo referir-se vários exemplos de funções  $\chi$  com tais requisitos. Um modo de escolher a função  $\chi$  resulta de usar alguma das funções  $\psi$  e  $\rho$  apresentados para o modelo de localização, definindo  $\chi$ , por exemplo, por  $\chi = \psi^2$  ou  $\chi = \rho$ . Assim, passam a apresentar-se algumas das possibilidades disponíveis na bibliografia (Pires e Branco, 2007).

- (i) Um estimador-M robusto para um modelo de escala normal, proposto por Huber, é definido por

$$\chi(x) = [\psi_c(x)]^2 = \begin{cases} x^2 & , |x| \leq c \\ c^2 & , |x| > c \end{cases},$$

ou por  $\psi(x) = \chi(x) - b$ , com  $b$  uma constante fixada de modo a que o estimador seja consistente sob o modelo normal, isto é, faz-se  $b = \int \chi(u) d\Phi(u)$ , para que  $\int \psi(u) d\Phi(u) = 0$ . Pode ainda deduzir-se desta última, a seguinte expressão para  $b$ , onde  $\Phi$  representa a função distribuição da distribuição normal padrão

$$b(c) = F(c^2, \chi_3^2) + 2c^2\Phi(-c), \quad (4.2.16)$$

e  $F(a, \chi_\nu^2)$  representa o valor da função de distribuição de um qui-quadrado com  $\nu$  graus de liberdade, no ponto  $a$ .

- (ii) Também é frequente usar a função  $\rho$  definida para o estimador bponderado de



Tukey para o modelo de localização, e que, para o parâmetro de escala, corresponde à função  $\chi$  seguinte,

$$\chi(x) = \begin{cases} \left(\frac{x}{r}\right)^6 - 3\left(\frac{x}{r}\right)^4 + 3\left(\frac{x}{r}\right)^2, & |x| \leq r \\ 1, & |x| > r \end{cases},$$

na forma normalizada, em que  $0 \leq \chi \leq 1$ , equivalente à anterior.

De entre os estimadores-M de escala aqui referidos, o estimador de Huber é o que apresenta melhores propriedades, apesar da grande dificuldade que existe em aliar eficiência e robustez na estimação do parâmetro de escala.

De modo a garantir o máximo ponto de ruptura ou valores elevados da eficiência assintótica para os correspondentes estimadores, sob o modelo normal, encontram-se disponíveis na bibliografia valores de constantes de afinação correspondentes às funções  $\chi$ , atrás referidas.

Na prática, acontece frequentemente que o parâmetro de localização é desconhecido, mesmo quando o interesse principal é estimar o parâmetro de escala. Nesses casos, é necessário determinar previamente uma estimativa de localização,  $T_n(x_1, \dots, x_n)$ , e depois a estimativa de escala para as observações centradas  $x_i^* = x_i - T_n$ . Deste modo, o estimador de escala resultante é invariante em relação à localização, de acordo com a condição (4.2.2), e equivariante em relação à escala, de acordo com a condição (4.2.3), desde que o estimador  $T_n$  usado seja equivariante em relação à localização e à escala (de acordo com a **Definição 4.2.3**).

O estimador de localização auxiliar  $T_n$  deve ser robusto para não afectar as propriedades de robustez do estimador de escala, sendo usualmente recomendada a mediana das observações.

De seguida, apresenta-se a definição que generaliza o conceito de estimador-M ao caso em que o parâmetro  $\theta$  é multivariado, com  $p$  componentes; as observações podem referir-se ao caso uni ou multivariado.

**Definição 4.2.8.** Um **estimador-M** é um estimador  $\mathbf{T}_n$  que seja solução do problema

$$\min \sum_{i=1}^n \rho(\mathbf{X}_i, \mathbf{T}_n), \quad (4.2.17)$$

onde  $\rho$  é uma função real arbitrária. Doutro modo, pode dizer-se que  $\mathbf{T}_n$  é um estimador-M se for solução da equação vectorial

$$\sum_{i=1}^n \psi(\mathbf{X}_i, \mathbf{T}_n) = \mathbf{0}. \quad (4.2.18)$$

□

Se  $\rho$  for diferenciável, tem-se que (4.2.17) é equivalente a (4.2.18), para  $\psi(\mathbf{x}, \boldsymbol{\theta}) = \frac{\partial \rho(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ .

A função de influência do funcional  $\mathbf{T}$ , equivalente ao estimador  $\mathbf{T}_n$ , definido por

$$\int \psi(\mathbf{x}, \mathbf{T}(F)) d\mathbf{F}(\mathbf{x}) = \mathbf{0},$$

pode ser deduzida, como no caso unidimensional, resultando em

$$IF(\mathbf{x}, \mathbf{T}, F) = \mathbf{M}(\psi, F)^{-1} \psi(\mathbf{x}, \mathbf{T}(F)),$$

onde  $\mathbf{M}$  é uma matriz  $(p \times p)$  definida por

$$\mathbf{M}(\psi, F) = - \int \left[ \frac{\partial \psi(\mathbf{y}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]_{\boldsymbol{\theta}=\mathbf{T}(F)} dF(\mathbf{y}).$$

De modo idêntico ao que acontece no caso univariado, a  $IF$  dos estimadores-M verifica as seguintes propriedades no caso multivariado:

1. A função de influência tem esperança nula, isto é

$$E_F[IF(\mathbf{x}, \mathbf{T}, F)] = \mathbf{0}.$$

2. Sob condições de regularidade, tem-se que

$$E_F[IF(\mathbf{x}, \mathbf{T}, F) IF(\mathbf{x}, \mathbf{T}, F)'] = V(\mathbf{T}, F),$$

onde  $V(\mathbf{T}, F)$  é a variância assintótica do estimador  $T$ .

Particularizando para  $p = 2$ , no caso de  $\psi$  ser simétrica, com componentes  $\psi_1$  e  $\psi_2$ , usam-se  $\psi_1$  ímpar e  $\psi_2$  par e podem apresentar-se alguns exemplos:

- Para o modelo normal, a média e o desvio padrão amostrais, para estimar os parâmetros  $\mu$  e  $\sigma$ , correspondem aos estimadores-M definidos por

$$\psi(z) = \begin{pmatrix} z & z^2 - 1 \end{pmatrix}'.$$

- Outro par de estimadores, consistentes para os parâmetros do modelo normal, é formado pela mediana e pelo desvio absoluto mediano amostrais, que correspondem aos estimadores-M definidos por

$$\psi(z) = \begin{pmatrix} \text{sinal}(z) & \text{sinal}(|z| - \Phi^{-1}(3/4)) \end{pmatrix}'.$$

- O estimador de Huber de localização e de escala é definido por

$$\psi(z) = \begin{pmatrix} \psi_b(z) & \psi_c(z)^2 - b(c) \end{pmatrix}'. \text{ Neste caso, as estimativas são obtidas por processos iterativos, sendo usual considerar a simplificação resultante de fazer } b = c. \text{ Este estimador tem ponto de ruptura global dado por } \varepsilon^* = \frac{b(c)}{b(c) + c^2}, \text{ quer para a componente de localização quer para a de escala, sendo o valor } b(c) \text{ o definido em (4.2.16). Como este valor de } \varepsilon^* \text{ é inferior aos pontos de ruptura dos estimadores-M de localização e de escala, quando considerados separadamente, alguns autores sugerem que se estimem separadamente os parâmetros de localização e de escala.}$$

### 4.2.3 Outras classes de estimadores

Embora os estimadores-M sejam a base da estimação robusta, foram sendo investigadas outras famílias de estimadores que se passam a referir. Dar-se-á um desenvolvimento especial à apresentação do estimador Ortogonalizado de *Gnanadesikan-Kettenring* (*OGK* de *Orthogonalized Gnanadesikan-Kettenring*), uma vez que este estimador será utilizado no capítulo seguinte.

#### Estimadores-L

Os estimadores-L são definidos por combinações lineares de estatísticas de ordem. A mediana é o exemplo mais comum para o parâmetro de localização e a amplitude inter-quartis para o parâmetro de escala. Este tipo de estimadores nem sempre constitui uma alternativa aos estimadores-M, pois a sua aplicação a casos multivariados não é directa (Fraiman *et al.*, 1999); por outro lado, as suas propriedades nem sempre os tornam preferenciais.

## Estimadores-R

Os estimadores-R baseiam-se em postos (*ranks*). Como estimador do parâmetro de localização, destaca-se o estimador de Hodges-Lehamnn (Hodges e Lehman, 1963) definido pela mediana das médias de todos os pares de observações. O estimador-R do parâmetro de escala mais divulgado é o estimador  $Q_n$ , proposto por Rousseeuw e Croux (1993). Este estimador é um dos que melhor consegue conciliar robustez com eficiência, em modelos de escala normais. O estimador  $Q_n$  corresponde ao primeiro quartil da amostra formada pelos valores absolutos das diferenças entre todos os pares de observações. Estes estimadores apresentam dificuldades na aplicação aos casos multivariados, tal como acontecia com os anteriores.

## Estimadores-W

Estes estimadores foram já apresentados na secção anterior e resultam de utilizar estimadores usuais, como a média ou a variância, depois de ponderados.

## Estimadores- $\tau$

Estes estimadores foram propostos por Yohai e Zamar (1988) como forma de ultrapassar a dificuldade de conseguir um elevado ponto de ruptura e uma eficiência elevada para os estimadores-M de escala. A ideia base consiste em começar por usar um estimador-M, com ponto de ruptura máximo e depois, usando este como ponto de partida, calcular um estimador-M com elevada eficiência. Deste modo, consegue-se um estimador com elevado ponto de ruptura, como o primeiro, e elevada eficiência, como o segundo. Estes estimadores podem também aplicar-se ao caso multivariado (Lopuhaä, 1991).

## Estimador *OGK*

O estimador robusto *OGK* foi proposto por Maronna e Zamar (2002). O estimador resulta de uma proposta anterior de Gnanadesikan e Kettenring (1972), para estimar a covariância entre duas variáveis aleatórias  $X$  e  $Y$ , com base na relação

$$\sigma_{XY} = \frac{\sigma^2(X + Y) - \sigma^2(X - Y)}{4}, \quad (4.2.19)$$

onde  $\sigma^2(\cdot)$  é uma representação funcional da variância. Para tornar robusto o estimador de  $\sigma_{XY}$ , deve ser usado um funcional robusto  $\sigma$ ; Gnanadesikan e Kettenring usaram o desvio padrão aparado.

Havendo mais do que duas variáveis, a respectiva matriz de covariâncias pode ser estimada de forma robusta, estimando componente a componente, formando a matriz estimada  $\hat{\Sigma}_{GK}$ . A matriz  $\hat{\Sigma}_{GK}$  é simétrica, mas nem sempre é semidefinida positiva. Para ultrapassar esta dificuldade, Maronna e Zamar (2002) usaram uma transformação de  $\hat{\Sigma}_{GK}$ , de modo a torná-la uma matriz ortogonal. Para resumir a ideia, recorde-se que para um vector aleatório  $\mathbf{x} \in \mathbb{R}^p$  e  $\mathbf{a} \in \mathbb{R}^p$ , se tem

$$\sigma^2(\mathbf{a}'\mathbf{x}) = \mathbf{a}'\sigma^2(\mathbf{x})\mathbf{a},$$

onde  $\sigma$  é um funcional de escala.

O estimador *OGK* consiste num processo de estimação que envolve estimadores robustos, aproximadamente equivariantes, de localização e de escala, conduzindo sempre a uma matriz de estimativas semidefinida positiva.

Uma vez que se optou por utilizar este procedimento nos capítulos seguintes, justifica-se apresentar o algoritmo de cálculo das estimativas *OGK*. No entanto, do ponto de vista da sua aplicação, é desnecessário programar este algoritmo, uma vez que o estimador está já disponível no programa *R*.

Para uma matriz  $(n \times p)$  de observações,  $\mathbf{X} = [x_{ij}]$ , representando por  $\mathbf{x}'_i$  as linhas e por  $\mathbf{x}_j$  as colunas,  $i = 1, \dots, n$  e  $j = 1, \dots, p$ , podem definir-se os estimadores *OGK* de localização  $\hat{\boldsymbol{\mu}}(\mathbf{X})$  e de escala  $\hat{\Sigma}(\mathbf{X})$  de acordo com os seguintes passos:

1. Calcular uma matriz de dados normalizados  $\mathbf{Y}$  com colunas  $\mathbf{y}_j = \frac{\mathbf{x}_j}{\hat{\sigma}(\mathbf{x}_j)}$  e com linhas  $\mathbf{y}'_i$ , onde

$$\mathbf{y}_i = \mathbf{D}^{-1}\mathbf{x}_i, \text{ onde } \mathbf{D} = \text{diag}(\hat{\sigma}(\mathbf{x}_j)), j = 1, \dots, p, i = 1, \dots, n$$

e  $\hat{\sigma}(\cdot)$  representa uma medida de dispersão robusta univariada.

2. Calcular a estimativa robusta da matriz de correlações de  $\mathbf{X}$ , denotada por  $\mathbf{U} = [U_{jk}]$ , considerando-a como matriz de covariâncias de  $\mathbf{Y}$ , aplicando a relação (4.2.19) às colunas  $\mathbf{y}_j$ , de acordo com

$$U_{jj} = 1 \text{ e } U_{jk} = \frac{\hat{\sigma}^2(\mathbf{y}_j + \mathbf{y}_k) - \hat{\sigma}^2(\mathbf{y}_j - \mathbf{y}_k)}{4}, j \neq k.$$

3. Calcular os valores próprios  $\lambda_j$  e os vectores próprios  $\mathbf{e}_j$  da matriz  $\mathbf{U}$  e definir a matriz  $\mathbf{E}$ , cujas colunas são os vectores próprios,  $\mathbf{e}_j$ . De seguida, utilizar a decomposição em componentes principais de  $\mathbf{Y}$ ,  $\mathbf{U} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}'$ , onde  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p)$ ,  $\lambda_j \geq 0$ .
4. Calcular a matriz  $\mathbf{Z}$  com linhas  $\mathbf{z}_i'$  tais que

$$\mathbf{z}_i = \mathbf{E}'\mathbf{y}_i = \mathbf{E}'\mathbf{D}^{-1}\mathbf{x}_i, i = 1, \dots, n.$$

As componentes principais de  $\mathbf{Y}$  são  $(\mathbf{z}_1, \dots, \mathbf{z}_p)$ , ou seja, as colunas da matriz  $\mathbf{Z}$ .

5. Calcular  $\hat{\sigma}(\mathbf{z}_j)$  e  $\hat{\mu}(\mathbf{z}_j)$ ,  $j = 1, \dots, p$ , onde  $\hat{\mu}(\cdot)$  representa uma medida de localização robusta univariada, e formar

$$\mathbf{\Gamma} = \text{diag}[\hat{\sigma}^2(\mathbf{z}_1) \dots \hat{\sigma}^2(\mathbf{z}_p)] \text{ e } \boldsymbol{\nu} = \begin{pmatrix} \hat{\mu}(\mathbf{z}_1) & \dots & \hat{\mu}(\mathbf{z}_p) \end{pmatrix}'.$$

Note-se que, deste modo, as componentes de  $\mathbf{\Gamma}$  são sempre não negativas.

6. Calcular  $\mathbf{X}$ , de acordo com

$$\mathbf{x}_i = \mathbf{A}\mathbf{z}_i \text{ e } \mathbf{A} = \mathbf{D}\mathbf{E},$$

e obter

$$\hat{\boldsymbol{\Sigma}}(\mathbf{X}) = \mathbf{A}\mathbf{\Gamma}\mathbf{A}' \text{ e } \hat{\boldsymbol{\mu}}(\mathbf{X}) = \mathbf{A}\boldsymbol{\nu}.$$

Os autores sugerem que se usem a mediana e o  $MAD$  como valores iniciais para  $\hat{\mu}_0$  e  $\hat{\sigma}_0$ , respectivamente, e que  $\hat{\mu}$  e  $\hat{\sigma}$  sejam obtidos por

$$\hat{\mu} = \frac{\sum_{i=1}^n x_i w_i}{\sum_{i=1}^n w_i} \text{ e } \hat{\sigma}^2 = \frac{\sigma_0^2}{n} \sum_{i=1}^n \rho\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}_0}\right),$$

com  $w_i = W\left(\frac{x_i - \hat{\mu}_0}{\hat{\sigma}_0}\right)$ , sendo apresentadas propostas para as funções  $W$  e  $\rho$ , de forma a conseguir um bom equilíbrio entre robustez e eficiência.

Maronna e Zamar (2002) recomendam que o procedimento comece por  $\hat{\boldsymbol{\mu}}_{(0)} = \hat{\boldsymbol{\mu}}(\mathbf{X})$  e  $\hat{\boldsymbol{\Sigma}}_{(0)} = \hat{\boldsymbol{\Sigma}}(\mathbf{X})$  e que se calculem iterativamente

$$\hat{\boldsymbol{\Sigma}}_{(k+1)}(\mathbf{X}) = \mathbf{A}\hat{\boldsymbol{\Sigma}}(\mathbf{Z}_{(k)})\mathbf{A}' \text{ e } \hat{\boldsymbol{\mu}}_{(k+1)}(\mathbf{X}) = \mathbf{A}\hat{\boldsymbol{\mu}}(\mathbf{Z}_{(k)}),$$

onde  $\mathbf{Z}_{(k)}$  é a matriz cujas linhas são  $\mathbf{z}'_i = (\mathbf{A}^{-1}\mathbf{x}_i)'$ . Recomendam apenas duas iterações.

Como se referiu, o cálculo das estimativas pode ser feito recorrendo à função *CovOgk* disponível na *package rrcov* do programa R.

Os processos que se vieram a referir podem ser usados para estimar os elementos da matriz de covariâncias. Tendo em consideração que esta é uma medida de dispersão multivariada, existem actualmente outras propostas que têm recolhido a preferência dos investigadores, pelos bons resultados a que têm conduzido. A ideia base deste tipo de estimadores é idêntica à que se usa no estimador dos parâmetros de um modelo de regressão, quando se escolhem os valores que minimizam os resíduos  $e_i$ . Neste caso, vão obter-se estimativas multivariadas de localização e de dispersão de forma a minimizar distâncias  $d_i$ . Para uma amostra  $\mathbf{X}$ , designa-se por  $\mathbf{d}(\mathbf{X}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$  o vector com componentes  $d(\mathbf{x}_i, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ ,  $i = 1, \dots, n$ , e seja  $\hat{\sigma}$  um estimador robusto da dispersão. Definem-se os estimadores  $\hat{\boldsymbol{\mu}}$  e  $\hat{\boldsymbol{\Sigma}}$  como solução do problema

$$\min \left\{ \hat{\sigma} \left[ \mathbf{d}(\mathbf{X}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) \right] \right\}, \quad (4.2.20)$$

com  $\hat{\boldsymbol{\mu}} \in R^p$  e  $\hat{\boldsymbol{\Sigma}}$  uma matriz  $p \times p$  simétrica e semidefinida positiva tal que  $|\hat{\boldsymbol{\Sigma}}| = 1$ . Os estimadores assim definidos são equivariantes.

### **Estimador do elipsóide de volume mínimo ou *MVE* (de *Minimum Volume Ellipsoid*)**

Os estimadores *MVE* de localização e de escala multivariados foram introduzidos por Rousseeuw (1985). Estes estimadores são definidos como solução de um problema do tipo de (4.2.20), em que se considera  $\hat{\sigma}$  a mediana amostral e se minimiza a mediana dos quadrados das distâncias (ideia idêntica à do estimador *LMS* que será apresentado na **Secção 4.3**). Deste modo, entre todos os elipsóides  $\{\mathbf{x} : \mathbf{d}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \leq 1\}$  que contêm pelo menos metade dos pontos da amostra, o estimador *MVE* corresponde ao que tem volume mínimo. Geometricamente, o estimador *MVE* de localização corresponde ao centro desse elipsóide e o estimador *MVE* de escala corresponde à matriz que define a forma do elipsóide. Na prática, a utilização destes estimadores pode ser feita recorrendo à função *covMVE* disponível na *package rrcov* do R.

O estimador *MVE* tem elevado ponto de ruptura, mas tem eficiência muito baixa.

### **Estimador da matriz de covariâncias de determinante mínimo ou *MCD* (de *Minimum Covariance Determinant*)**

Os estimadores *MCD* de localização e de escala multivariados foram igualmente propostos por Rousseeuw (1985). Também estes estimadores são definidos como solução de um problema do tipo de (4.2.20), mas considerando agora que  $\hat{\sigma}$  é a escala aparada e se minimiza a soma aparada dos quadrados das distâncias (ideia idêntica à do estimador *LTS*, igualmente apresentado na **Secção 4.3**). Neste caso, para cada elipsóide  $\{\mathbf{x} : \mathbf{d}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \leq 1\}$ , contendo pelo menos  $h$  pontos da amostra, calcula-se a matriz de covariâncias do conjunto de observações contidas no elipsóide e o respectivo determinante. O estimador *MCD* corresponde ao elipsóide tal que o determinante da matriz de covariâncias é mínimo. O máximo ponto de ruptura para o estimador *MCD* é conseguido considerando  $h = n - \left\lfloor \frac{n-p}{2} \right\rfloor$ . Embora a obtenção das estimativas *MCD*, possa ser feita computacionalmente, recorrendo à função *covMcd* da *package robustbase* do *R*, o tempo necessário à obtenção das estimativas é muito elevado. Este aspecto suscitou o desenvolvimento de algoritmos mais eficientes, como o *FAST-MCD*, proposto por Rousseeuw e Van Driessen (1999), o qual se encontra igualmente disponível no programa *R*.

## **4.3 Regressão robusta**

Dada a semelhança entre os modelos de regressão multivariada e os *SEM*, esta secção é devotada à apresentação de técnicas da regressão robusta.

A estimação dos parâmetros de um modelo de regressão está inevitavelmente associada ao método dos mínimos quadrados. Quando a estatística robusta começou a desenvolver-se, o método dos mínimos quadrados começou a receber algumas críticas desfavoráveis. Efectivamente, o método dos mínimos quadrados não conduz a estimadores robustos, o que pode trazer consequências graves perante desvios das hipóteses, como sejam erros não normais ou existência de observações atípicas. Os métodos da regressão robusta propõem-se dar uma resposta mais eficaz, diminuindo o efeito do



afastamento das hipóteses do modelo, procurando por outro lado manter uma eficiência elevada quando as hipóteses são válidas.

Nos livros de regressão mais recentes a questão da robustez é já contemplada, quer pela existência de secções próprias, ou mesmo de capítulos dedicados ao tema - por exemplo, Ryan (1997) e Draper e Smith (1998) dedicam um capítulo à regressão robusta. No entanto, não se pode deixar de recomendar a leitura de Rousseeuw e Leroy (1987), onde se desenvolve exclusivamente este tema e sem exigir dos leitores grandes conhecimentos de matemática. De seguida, faz-se uma breve apresentação da estimação clássica em modelos de regressão, para passar a sumariar algumas alternativas robustas, referindo-se as propriedades mais relevantes de cada método.

### 4.3.1 Modelo de regressão e método dos mínimos quadrados

O modelo de regressão linear múltipla é definido por uma relação linear entre um conjunto de observações de uma variável resposta e  $p$  variáveis explicativas. Para uma amostra de dimensão  $n$ , pode caracterizar-se na forma matricial, por

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (4.3.1)$$

onde  $\mathbf{y} = (y_1, \dots, y_n)'$  é formado por  $n$  observações de uma variável resposta,  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_{p-1})'$  é um vector de parâmetros desconhecidos,  $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)'$  é o vector dos erros do modelos e as  $n$  observações dos  $p$  regressores estão organizadas na matriz

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p-1} \\ 1 & x_{21} & \dots & x_{2p-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \dots & x_{np-1} \end{bmatrix} = \begin{bmatrix} \mathbf{x}'_1 \\ \vdots \\ \mathbf{x}'_n \end{bmatrix}.$$

As hipótese usuais consistem em assumir que os erros  $\varepsilon_i$  são não correlacionados e identicamente distribuídos, com média nula e variância constante, para  $i = 1, \dots, n$ , e que a matriz  $\mathbf{X}$  é determinística (ainda que corresponda a realizações de variáveis aleatórias). Sob as hipótese assumidas, a equação de regressão é  $E(\mathbf{y}|\mathbf{X}) = \mathbf{X}\boldsymbol{\beta}$ .

O modelo de regressão linear multivariado (*MRLM*) generaliza o modelo anterior, no sentido em que considera mais do que uma variável dependente. Neste modelo, admite-se que cada variável resposta segue um modelo de regressão linear múltipla,

com as respectivas hipóteses estabelecidas e que os erros das diferentes equações são não correlacionados.

O *MRLM* com  $M$  variáveis resposta pode caracterizar-se na forma matricial, por

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

mas agora  $\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 & \dots & \mathbf{y}_M \end{bmatrix}$  é uma matriz  $(n \times M)$ , formada pelas  $n$  observações

das  $M$  variáveis resposta,  $\boldsymbol{\beta}$  uma matriz  $(p \times M)$  de parâmetros desconhecidos,  $\boldsymbol{\varepsilon}$  uma matriz  $(n \times M)$  de erros aleatórios e  $\mathbf{X}$  a matriz  $(n \times p)$  das observações dos regressores.

Ao encarar o *MRLM* como um sistema, assume-se que não há correlação entre as equações. Deste modo, a estimação do *MRLM* pode ser estudada através do estudo da estimação do modelo de regressão linear múltipla, que se verifica para cada uma das equações. Apresentam-se de seguida as principais técnicas de regressão robusta apenas para o modelo de regressão linear múltipla referido em (4.3.1).

Um dos principais objectivos no estudo do modelo de regressão consiste na estimação de  $\boldsymbol{\beta}$ . O método dos mínimos quadrados é o método mais intuitivo e mais utilizado. Baseia-se na determinação do estimador  $\hat{\boldsymbol{\beta}}$  que minimiza a distância dos pontos à curva de regressão. Assim,  $\hat{\boldsymbol{\beta}}$  é a solução do problema de minimização da função

$$Q(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

O estimador resultante é da forma

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}.$$

Supondo verdadeiras as hipóteses do modelo, o teorema de Gauss-Markov garante que o estimador de mínimos quadrados é centrado para  $\boldsymbol{\beta}$  e que têm variância mínima, de entre todos os estimadores centrados que sejam combinações lineares das respostas. No caso de as hipóteses exigidas não serem satisfeitas, estas propriedades desejáveis não ficam asseguradas, havendo o risco do método produzir resultados muito enganadores.

Quando se pretende estabelecer inferências sobre o modelo, em geral assume-se que os erros seguem distribuição normal. Nessas circunstâncias, as componentes de  $\boldsymbol{\varepsilon}$  são

*i.i.d.*, com  $\varepsilon \sim \mathbf{N}_n(\mathbf{0}, \sigma^2 \mathbf{I})$ , verificando-se que o estimador de máxima verosimilhança coincide com o estimador de mínimos quadrados.

Apesar da facilidade de utilização e das boas propriedades que goza o estimador dos mínimos quadrados, lamentavelmente não é um estimador robusto, uma vez que tem ponto de ruptura nulo e função de influência ilimitada. Significa isto que o método é extremamente sensível a desvios das hipóteses do modelo e que basta a existência de uma única observação atípica para conduzir a estimativas desastrosas.

O grande número de situações aplicadas que requerem a modelação por regressão, motivou o aparecimento e a investigação de diversos métodos robustos. Aliás, foi precisamente na área da regressão que surgiu a maior parte dos estudos sobre robustez.

Referem-se a seguir alguns dos métodos mais conhecidos de regressão robusta, os quais se propõem ajustar uma curva de regressão que traduza a contribuição da maioria dos dados.

### Estimador de mínimos desvios absolutos

O método dos mínimos desvios absolutos, ou método *LAD* (de *Least Absolute Deviations*), é semelhante ao método dos mínimos quadrados, uma vez que procura a regressão que melhor se ajusta ao conjunto de dados, mas difere dele, no sentido em que minimiza a soma dos valores absolutos dos resíduos, em vez da soma dos quadrados dos resíduos.

Deste modo, o estimador *LAD* é o valor de  $\beta$  que é solução do problema

$$\min \sum_{i=1}^n \left| y_i - \left( \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_{p-1} x_{ip-1} \right) \right| = \min \sum_{i=1}^n |e_i|.$$

O método surgiu várias décadas antes do método dos mínimos quadrados, tendo uma utilização muito mais reduzida devido, por um lado, a obrigar a cálculos mais morosos, por outro, a conduzir a soluções múltiplas, em geral definidas implicitamente e obtidas por métodos iterativos.

Porém, o método dos mínimos desvios absolutos é superior ao método dos mínimos quadrados, do ponto de vista da robustez, sendo no entanto muito sensível a observações atípicas em  $\mathbf{x}$ . Por este motivo, tem ponto de ruptura nulo. Por fim, é importante referir, que no caso de os erros serem normais, o *LAD* é pouco eficiente.

## Estimadores-M

Os **estimadores-M** dos coeficientes da regressão foram introduzidos por Huber (1973) e podem ser entendidos como uma extensão dos estimadores de mínimos quadrados. A ideia básica já foi explicada na **Subsecção 4.2.2** e consiste em substituir a função objectivo a minimizar - a soma dos quadrados dos resíduos - por uma outra função de resíduos, ou seja, em determinar  $\hat{\beta}$  como solução de um problema do tipo  $\min \sum_{i=1}^n \rho(e_i)$ , para funções  $\rho$  convenientes. Assim, o estimador de mínimos quadrados pode ser visto como um estimador-M em que  $\rho(u) = u^2$ ; o *LAD* é um estimador-M onde  $\rho(u) = |u|$ .

Mais uma vez, note-se que, sob condições de regularidade, resolver o problema

$$\min \sum_{i=1}^n \rho(e_i) = \min \sum_{i=1}^n \rho(y_i - \mathbf{x}_i' \hat{\beta}) \quad (4.3.2)$$

equivale a resolver o sistema de equações obtido pelas condições de primeira ordem

$$\begin{aligned} \sum_{i=1}^n \psi(y_i - \mathbf{x}_i' \hat{\beta}) \mathbf{x}_i &= \mathbf{0} \\ \Leftrightarrow \sum_{i=1}^n \psi(e_i) \mathbf{x}_i &= \mathbf{0}, \end{aligned} \quad (4.3.3)$$

onde  $\rho' = \psi$ .

Embora os estimadores de mínimos quadrados e o *LAD* sejam equivariantes em relação à escala, os estimadores-M com função  $\psi$  geral não verificam esta propriedade. Para que ela seja verificada é necessário substituir  $e_i$  por  $\frac{e_i}{\hat{\sigma}}$ , onde  $\hat{\sigma}$  é uma estimativa robusta do parâmetro de escala  $\sigma$  da distribuição dos erros  $\varepsilon_i$ . É habitual usar-se o desvio absoluto mediano ou *MAD*. Deste modo, a condição (4.3.3) é substituída por

$$\sum_{i=1}^n \psi\left(\frac{e_i}{\hat{\sigma}}\right) \mathbf{x}_i = \mathbf{0} \Leftrightarrow \sum_{i=1}^n \psi\left(\frac{y_i - \mathbf{x}_i' \hat{\beta}}{\hat{\sigma}}\right) \mathbf{x}_i = \mathbf{0}. \quad (4.3.4)$$

A escolha da função  $\psi$  é da maior importância para que o estimador-M que seja eficiente e robusto, seguindo as indicações dadas na **Subsecção 4.2.2**. Por exemplo, supondo a distribuição normal, se  $\psi$  for a função de Huber, o estimador resultante é mais eficiente do que o estimador *LAD* e é robusto em relação a observações atípicas em relação a  $y$ . No entanto, o mesmo não acontece em relação a observações atípicas em relação a  $\mathbf{x}$ . Apesar da desvantagem referida, estes estimadores continuam a ser

bastante importantes, na medida em que a sua base conceptual serviu de modelo para a criação de novos e melhores estimadores, como veremos no seguimento do trabalho.

Assim, para usar um estimador-M é necessário, por um lado, escolher a função  $\psi$  (foram já apresentadas algumas possibilidades na **Subsecção 4.2.2**); por outro, decidir os valores das constantes de afinação, de forma equilibrar o grau de eficiência com o ponto de ruptura. Encontram-se na bibliografia (por exemplo em Staudte e Sheather, 1990, Maronna *et al.*, 2006 ou Pires e Branco, 2007) tabelas de valores destas constantes para diferentes cenários.

### Estimadores-GM

Os estimadores-M generalizados (**estimadores-GM**) surgiram para melhorar os estimadores-M, pelo facto de não serem resistentes a observações atípicas em relação a  $\mathbf{x}$ . Nesse sentido pensou-se atribuir um peso  $w_i$  a cada  $\mathbf{x}_i$ , de forma a limitar a influência de  $\mathbf{x}_i$  sobre as estimativas, devendo o valor de  $w_i$  ser tanto mais pequeno, quanto mais influente for  $\mathbf{x}_i$ .

Estes estimadores também são designados por estimadores de influência limitada. De entre as famílias de estimadores-GM mais conhecidas, destacam-se os **estimadores de Mallows**, definidos como solução de uma equação do tipo

$$\sum_{i=1}^n w_i(\mathbf{x}_i) \psi\left(\frac{e_i}{\hat{\sigma}}\right) \mathbf{x}_i = \mathbf{0},$$

e os **estimadores de Schweppe**, que diferem dos anteriores pelo facto de atribuírem pesos não só aos  $\mathbf{x}_i$ , mas também aos resíduos  $e_i$ . Estes últimos estimadores são obtidos como solução de uma equação da forma

$$\sum_{i=1}^n w_i(\mathbf{x}_i) \psi\left(\frac{e_i}{w_i(\mathbf{x}_i) \hat{\sigma}}\right) \mathbf{x}_i = \mathbf{0}.$$

Os estimadores Schweppe apresentam boas propriedades (ver Maronna *et al.*, 2006), nomeadamente, para  $\psi$  limitada, têm função de influência limitada e têm ponto de ruptura positivo, podem atingir um grau de eficiência elevada na distribuição normal, mas são muito criticados pelo facto do seu ponto de ruptura decrescer quando aumenta o número de parâmetros,  $p$ . Isso significa que são estimadores adequados para modelos de regressão lineares simples ou com poucos regressores, mas menos recomendáveis noutras situações.

### Estimadores da mínima mediana dos quadrados

Enquanto que os estimadores até agora apresentados foram obtidos alterando a função objectivo que se pretende minimizar, mas mantendo o somatório para todas as observações, o estimador da mínima mediana dos quadrados ou *LMS* (de *Least Median of Squares*) propõe não só que se altere a função dos resíduos, mas também que o somatório seja substituído por um operador robusto como, por exemplo, a mediana. Deste modo, o estimador *LMS*, proposto por Rousseeuw (1984), é definido como solução do problema

$$\min [med_i (e_i)^2] .$$

O estimador *LMS* é robusto em relação a observações atípicas em relação a  $y$  e em  $\mathbf{x}$  e atinge o ponto de ruptura máximo de 50%. No entanto em Rousseeuw e Leroy (1987) prova-se que o *LMS* tem eficiência assintótica nula, sob o modelo normal, o que faz com que ele não deva ser utilizado na maior parte das circunstâncias.

### Estimadores dos mínimos quadrados aparados

Para fazer face à falta de eficiência do *LMS*, em Rousseeuw (1984), os autores propõem outro estimador, o estimador aparado de mínimos quadrados ou *LTS* (de *Least Trimmed Squares*). Este estimador resulta de minimizar a soma aparada dos quadrados dos resíduos; isto é,  $\hat{\beta}_{LTS}$  é solução do problema

$$\min \sum_{i=1}^h e_{(i)}^2,$$

onde  $e_{(1)}^2 \leq e_{(2)}^2 \leq \dots \leq e_{(n)}^2$  são os quadrados dos resíduos ordenados e o somatório não inclui os  $n - h$  maiores resíduos absolutos. O valor de  $h$ , que representa o número de observações centrais a incluir no somatório, vai determinar o ponto de ruptura do estimador, uma vez que as restantes observações são "aparadas", o que faz com que o estimador não dependa delas.

Para este estimador, o valor máximo do ponto de ruptura é atingido quando  $h = [n/2] + [(p + 1)/2]$ . Quando  $h = n$ , o estimador coincide com o estimador de mínimos quadrados, logo tem ponto de ruptura zero.

Os cálculos envolvidos na determinação das estimativas por *LTS* são computacionalmente pesados, no entanto, a função *ltsReg*, que está disponível na *package robustbase*,

em  $R$ , permite obter rapidamente as estimativas, a partir de um algoritmo desenvolvido por Rousseeuw e Van Drissen (1999).

Embora a eficiência do estimador  $LTS$  seja bastante superior à do  $LMS$ , a eficiência do estimador  $LTS$  é ainda relativamente baixa. Assim, o  $LTS$  constitui um excelente processo para obter soluções iniciais em estimadores obtidos por processos iterativos.

### Estimadores-S

Um **estimador-S** resulta de minimizar outro estimador de escala dos resíduos, tendo sido proposto por Rousseeuw e Yohai (1984). Assim, por envolver um estimador de escala (*scale*), foi designado por estimador-S e obtém-se como solução do problema

$$\min S \left( e_1 \left( \hat{\beta} \right), \dots, e_n \left( \hat{\beta} \right) \right),$$

onde  $S(e_1(\beta), \dots, e_n(\beta))$  é um estimador-M de escala e, portanto, solução da equação

$$\sum_{i=1}^n \chi \left( \frac{e_i}{S} \right) = nb,$$

onde  $b$  está definido em (4.2.14) e é a constante de consistência associada à função  $\chi$ , de acordo com (4.2.15).

Uma vez que um estimador-S é um estimador-M, mantém-se a necessidade, já referida, de escolher a função  $\chi$  e os valores das constantes de afinação, de forma a equilibrar a eficiência com a robustez.

### Estimadores-MM

Os **estimadores-MM** foram propostos por Yohai (1987) e são actualmente considerados a melhor opção de entre os estimadores robustos, uma vez que permitem conjugar um elevado ponto de ruptura com uma elevada eficiência, sob o modelo normal. Dos estimadores referidos até agora, não há nenhum que verifique simultaneamente estes dois requisitos. De facto, aproveitando o resumo feito em Pires e Branco (2007):

- O estimador de mínimos quadrados tem eficiência máxima, mas tem ponto de ruptura nulo.
- O estimador  $LAD$  é pouco eficiente e tem ponto de ruptura nulo (em relação a  $\mathbf{x}$ ).

- Os estimadores *LMS*, *LTS* e *S* têm baixa eficiência e elevado ponto de ruptura.
- Os estimadores-M têm elevada eficiência, para certas funções  $\psi$  (ou  $\rho$ ), mas têm ponto de ruptura nulo.
- Os estimadores-GM podem ter elevada eficiência, dependendo da distribuição, mas o seu ponto de ruptura é decrescente para zero quando o número de regressores aumenta.

O nome destes estimadores - estimadores-MM - resulta do facto destes estimadores serem construídos com base em dois estimadores-M, um de localização e outro de escala. Assim, estes estimadores são estimadores-M calculados a partir de estimativas iniciais convenientes. Como se viu anteriormente, um estimador-M de regressão é obtido como solução do problema

$$\min \sum_{i=1}^n \rho \left( \frac{e_i}{\hat{\sigma}} \right),$$

de acordo com (4.3.2) e (4.3.4), onde  $\hat{\sigma}$  é um estimador de escala, sendo usualmente recomendado o *MAD*. Exceptuando casos particulares, um mínimo absoluto é difícil de conseguir, mas Yohai (1987) provou que se for encontrado um "bom" mínimo local, é possível conseguir um elevado ponto de ruptura e uma alta eficiência, escolhendo adequadamente a função  $\rho$ , que deve ser limitada. Um "bom" mínimo local pode ser obtido por aplicação do método *IRWLS*, definido em (4.2.11), a uma estimativa inicial adequada, sendo sugerida uma obtida usando o estimador-S. Repare-se que o estimador-MM resulta então de usar para  $\hat{\sigma}$  um estimador-M de escala e a garantir previamente que o ponto de ruptura é elevado, ou seja, é o estimador que corresponde ao mínimo local da função objectivo de um estimador-M que está mais próximo do estimador-S inicial.

Note-se que  $\hat{\sigma}$  estima o valor de  $\sigma$ , que representa o desvio padrão dos erros do modelo de regressão no caso homocedástico e que, portanto, é constante.

Nos programas R e S-plus encontram-se disponíveis funções que permitem efectuar os cálculos necessários à aplicação dos estimadores-MM.



## 4.4 Estimação robusta em *SEM*

Nesta secção, referem-se os principais trabalhos publicados sobre estimação robusta em *SEM*. Embora a estimação clássica neste tipo de modelos não seja robusta, uma vez que pode sofrer grandes perturbações quando se verificam desvios em relação ao modelo especificado ou quando existem observações atípicas entre os dados, alguns autores têm desenvolvido algumas propostas de estimadores robustos.

Como os estimadores mais usados em *SEM* resultam da aplicação do método dos mínimos quadrados, um procedimento que esteve na origem de vários estimadores robustos para estes modelos consiste em substituir a regressão de mínimos quadrados por métodos de regressão robusta. Assim, uma proposta (Amemiya, 1982) consistiu em desenvolver uma versão robusta do *2SLS*, substituindo o estimador *OLS* por um estimador robusto, em uma ou em ambas as fases. Amemiya (1982) apresentou dois estimadores que resultaram de aplicar o estimador *LAD* na segunda fase, ou em ambas as fases. Ambos os estimadores são mais eficientes do que o *2SLS* perante erros não normais, mas são menos eficientes sob a distribuição normal e não são robustos a observações atípicas em  $x$ . Krasker e Welsh (1985) propuseram um estimador ponderado de variáveis instrumentais, que consiste em usar um estimador-GM, ou de influência limitada, na segunda fase para estimar os parâmetros de cada equação. Posteriormente, Krasker (1986) avançou com um complemento desta proposta, usando um estimador-GM em ambas as fases. Este estimador é mais robusto do que os anteriores, mas continua a sofrer dos problemas já referidos para os estimadores-GM, ou seja, tem ponto de ruptura que tende para zero com o número de regressores.

Outra via usada para obter versões robustas de estimadores em *SEM*, consistiu em partir do estimador de máxima verosimilhança. Prucha e Kelejian (1984) apresentaram uma versão robusta do *MLIC*, para o caso em que a distribuição dos erros da forma reduzida é  $t$  de Student multivariada. Esta proposta permite obter estimadores que são robustos em relação a observações atípicas em  $y$ , mas não em relação a observações atípicas em  $x$ . Uma extensão do estimador de máxima verosimilhança para a forma reduzida e para a forma estrutural de um *SEM*, foram propostas por

Krishnakumar e Ronchetti (1997), onde é apresentada uma classe de estimadores robustos para estes modelos. Os estimadores propostos foram obtidos como solução da equação de verosimilhança generalizada e são resistentes a desvios do modelo normal e a observações atípicas, permitindo assegurar um melhor compromisso entre eficiência e robustez. Outra proposta (Kim e Muller, 2007) consistiu num estimador dos parâmetros estruturais de um *SEM*, que combina robustez com eficiência. Este estimador resulta da aplicação do estimador de Huber nas duas fases do processo de estimação, efectuando uma correcção de escala, ao usar previamente o estimador *MAD*. Conforme se referiu, os estimadores-M têm a desvantagem de ter ponto de ruptura nulo.

Tendo em consideração que o *SEM* na forma  $\mathbf{y} = \mathbf{Z}\boldsymbol{\delta} + \mathbf{e}$ , referida em (3.4.2), pode ser considerado um modelo de regressão com erros aparentemente não correlacionados, ou modelo *SUR* (do inglês *Seemingly Unrelated Regression*), importa também referir os estimadores robustos que têm sido propostos para estes modelos. O trabalho de Koenker e Portnoy (1990) inclui uma adaptação do estimador-M para o modelo *SUR*. No entanto, este estimador tem baixa eficiência se as variáveis dependentes estiverem correlacionados e tem ponto de ruptura nulo. Bilodeau e Duchesne (2000) desenvolveram uma versão do estimador-S adequada aos modelos *SUR*, que conduziu a um estimador robusto que mostraram ser mais eficiente do que os outros estimadores usados na comparação (máxima verosimilhança e estimador-M, proposto por Koenker e Portnoy (1990)). Este estimador vai ter em comum com os estimadores-S um ponto de ruptura elevado, mas baixa eficiência.

Outro trabalho que importa referir é o de Maronna e Yohai (1997), onde os autores estabeleceram três estratégias para obter estimadores robustos em *SEM*, com elevada eficiência sob o modelo normal e com ponto de ruptura elevado. A primeira proposta consiste em usar procedimentos robustos nas três fases do *3SLS*; na segunda opção, a sugestão passa pela transformação do *MLIC* num estimador robusto através da minimização do determinante da matriz de covariâncias dos resíduos e na última possibilidade é indicado um estimador que resulta da generalização do estimador- $\tau$  multivariado.

Sakata (2007) propôs um estimador que se baseia num princípio idêntico ao do estimador de variáveis instrumentais, isto é, parte da condição de ortogonalidade que

estabelece que os erros estruturais e as variáveis instrumentais são não correlacionados. Esta condição de ortogonalidade surge muitas vezes como consequência da restrição da média condicionada, que determina que a média dos erros estruturais condicionados pelas variáveis instrumentais é zero. A proposta deste autor consiste em usar antes a restrição da mediana condicionada.

Outra forma de obter estimadores robustos, consistiu em generalizar a ideia de estimação quantílica, proposta por Koenker e Basset (1978), ao modelo de equações simultâneas. Este processo de estimação é robusto em relação a erros não normais ou à presença de observações atípicas em relação a  $y$ , mas é sensível a observações atípicas relativas às variáveis independentes (Buchinsky, 1997). Chen e Portnoy (1996) aplicaram a regressão quantílica para desenvolver um estimador aparado de mínimos quadrados (*LTS*) em duas fases. Posteriormente, Chen *et al.* (2000a) apresentaram um estimador aparado de mínimos quadrados (*LTS*) em duas fases de tipo simétrico, que mostraram ser mais eficiente do que outras propostas robustas e, noutra proposta, os autores aplicaram regressão quantílica, de influência limitada do tipo Mallow's, para estimar os parâmetros estruturais de um *SEM* (Chen *et al.*, 2000b). Posteriormente, Chen *et al.* (2001) generalizaram aos *SEM* o estimador da média aparada, proposto por Welsh (1987) para o modelo de regressão linear. Um estimador de regressão quantílica em duas fases foi desenvolvido por Kim e Muller (2004), em que a segunda fase se baseia numa regressão quantílica com o mesmo quantil que é usado na regressão quantílica da primeira fase do processo de estimação. Kim e Muller (2004) mostraram que nestas condições fica assegurada a robustez do estimador proposto.

No entanto, os estimadores de *GMM* para *SEM* não tinham sido ainda estudados numa perspectiva de robustez, mas foi considerado haver todo o interesse em desenvolver uma versão robusta do *GMM*.

# Capítulo 5

## Versão robusta do estimador $GMM$

### 5.1 Introdução

Tradicionalmente a estimação dos parâmetros em modelos de equações simultâneas faz-se recorrendo ao princípio dos mínimos quadrados, supondo o modelo homocedástico. Basicamente, quando se opta por um método de informação limitada, o processo para obter o estimador  $2SLS$  consiste em duas fases; ao optar por um método informação completa, o processo de estimação é composto por três fases, originando o estimador  $3SLS$ . A alternativa robusta aos estimadores de mínimos quadrados foi conseguida pela "robustificação" de cada uma das fases do processo de estimação, ou seja, em cada fase substituiu-se a presença de um estimador tradicional por um estimador robusto.

Como se referiu em secções anteriores, o estimador  $GMM$  tem boas propriedades mas, no caso geral, não é robusto (ver Ronchetti e Trojani, 2001) e não foi encontrada na bibliografia investigada nenhuma versão robusta do estimador  $GMM$  em  $SEM$  até ao desenvolvimento da presente tese.

O presente capítulo tem como objectivo apresentar uma proposta de uma versão robusta do estimador  $GMM$ . Tal como os estimadores de mínimos quadrados, o estimador  $GMM$  também resulta de um processo composto por etapas. Para construir a alternativa robusta, em cada etapa do cálculo das estimativas por  $GMM$ , substituíram-se os procedimentos tradicionais por métodos robustos.

A investigação foi conduzida considerando modelos heterocedásticos e métodos de informação completa. Ou seja, optou-se por construir uma versão robusta para a forma

mais geral da matriz de covariâncias, estimando simultaneamente todas as equações do sistema. Apesar desta via exigir estudos mais complexos e mais trabalhosos, esta opção pareceu-nos justificável tendo em conta quatro razões: elegância matemática, possibilidade de particularização em modelos homocedásticos, superioridade dos *MIC*, teoricamente comprovada no presente trabalho e, finalmente, acessibilidade informática que não envolve um esforço computacional significativamente mais relevante para o utilizador.

Inspirada no artigo de Maronna e Yohai (1997), a investigação foi feita aplicando procedimentos robustos nos três passos do processo para a obtenção do estimador:

- (i) obtenção de estimativas iniciais dos resíduos,
- (ii) estimação da matriz de covariâncias  $\Omega$ ,
- (iii) resolução do problema de minimização.

Após a estimação dos parâmetros pelo método que se propõe, foram calculados os resíduos finais do processo. Como esses resíduos são multivariados, a qualidade das estimativas é avaliada através das médias das normas dos resíduos e dos números de condição (logaritmo de) da sua matriz de covariâncias amostral. Estes indicadores são sugeridos em Maronna *et al.* (2006) para situações multivariadas.

A inexistência de *software* disponível para o método de estimação que se propõe, exigiu que fossem desenvolvidos os correspondentes algoritmo e programa computacional. Tendo em conta, por um lado, o acesso livre ao programa *R* e, por outro lado, as funções estatísticas já programadas em *packages* adicionais, a componente computacional do trabalho foi efectuada em *R*. No entanto, houve a necessidade de construir originalmente a parte mais relevante do programa, pois a complexidade do modelo impede a utilização directa de muitos comandos pré-definidos, particularmente no que toca à robustez.

Com efeito, a *package systemfit* proporciona o cálculo de estimativas em modelos *SEM* por *3SLS* e por *GMM*, mas apenas sob a hipótese de homocedasticidade dos modelos. Além disso, os estimadores aí contemplados são os tradicionais, não robustos. Por isso, essa *package* serviu apenas de apoio numa fase inicial do estudo, tendo em vista a confirmação dos resultados obtidos por outras vias.

Por outro lado, as *packages*/comandos direccionadas para a estimação robusta, não foram concebidas para modelos com as características dos *SEM*; para além disso, embora no desenvolvimento do trabalho se tenha seguido um algoritmo inspirado no da regressão robusta, as funções disponíveis para a regressão robusta no programa *R* não podem ser directamente utilizadas, por motivos que decorrem da conjugação da interdependência entre equações e heterocedasticidade.

A tarefa adicional de programar a obtenção das estimativas, embora tenha exigido um esforço não desejado e um consumo de tempo significativo, teve o mérito de realçar algumas dificuldades associadas à especificidade dos *SEM*, as quais conduziram a novas reflexões de natureza estatística e consequentes modificações dos algoritmos inicialmente pensados. A título de exemplo, note-se que a heterocedasticidade motivou cuidados especiais com a simulação de amostras contaminadas, assim como implicou o recurso a técnicas de reamostragem, para possibilitar a utilização de estimadores robustos das matrizes de covariâncias.

Nas secções seguintes explica-se a metodologia robusta usada em cada uma das fases. Na última secção apresenta-se o algoritmo de cálculo para o caso geral (modelo heterocedástico e estimação por *MIC*) e a particularização para o caso homocedástico.

## 5.2 Obtenção de estimativas iniciais dos resíduos

O processo tradicional na determinação de estimativas iniciais consiste em estimar os coeficientes do sistema, separadamente para cada equação, ou seja, obter estimativas iniciais dos parâmetros por um método de informação limitada. Neste ponto, Greene (2003) sugere que se use o método dos mínimos quadrados com informação limitada, o que corresponde a obter as estimativas iniciais por *2SLS*, por intermédio de  $\hat{\delta}_{i,2SLS} = (\hat{\mathbf{Z}}_i' \hat{\mathbf{Z}}_i)^{-1} \hat{\mathbf{Z}}_i' \mathbf{y}_i$ , com  $\hat{\mathbf{Z}}_i = \begin{bmatrix} \hat{\mathbf{Y}}_i & \mathbf{X}_i \end{bmatrix}$ , já referido em (3.3.27). Contudo, Mittelhammer *et al.* (2000) recomenda que se obtenham as estimativas iniciais aplicando o método dos momentos generalizado a cada equação do sistema, ou seja, recorrendo ao estimador *GMMIL*, através de

$\hat{\delta}_{i,GMM} = [\mathbf{Z}_i' \mathbf{X} (\mathbf{X}' \mathbf{\Omega}_i \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z}_i]^{-1} \mathbf{Z}_i' \mathbf{X} (\mathbf{X}' \mathbf{\Omega}_i \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}_i$ , referido anteriormente em (3.7.4). Ambas as possibilidades conduzem a estimativas iniciais não robustas e que

podem ser muito enganadoras se houver observações atípicas na amostra. De facto, mesmo que a matriz  $\mathbf{\Omega}_i$  seja estimada de forma robusta, tanto o  $\hat{\boldsymbol{\delta}}_{i,2SLS}$  como  $\hat{\boldsymbol{\delta}}_{i,GMM}$  dependem directamente das observações através de  $\mathbf{Z}_i$ ,  $\mathbf{X}$  e  $\mathbf{y}_i$ , o que torna evidente que qualquer contaminação da amostra se pode reflectir sobre as estimativas produzidas.

Para conseguir boas estimativas iniciais, optou-se por recorrer a um método de regressão robusta, tendo sido escolhido o estimador *LTS* (apresentado na **Subsecção 4.3.2**), por ser um método recomendado para a obtenção da solução inicial, por ter ponto de ruptura não nulo e alguma eficiência sob o modelo normal. Deste modo, cada equação  $i$  do modelo foi considerada como uma equação de regressão múltipla e estimaram-se os respectivos parâmetros por *LTS*, ou seja, minimizando  $\sum_{l=1}^{h_i} e_{(l)}^2$ , onde  $e_{(l)}^2$  é o quadrado do resíduo de ordem  $l$ , isto é, não incluindo no somatório os  $T - h_i$  maiores resíduos absolutos, para  $i = 1, \dots, M$ . Nesse aspecto, consideram-se as configurações de base do programa.

Aplicando o estimador *LTS* à equação  $i$ , obteve-se  $\hat{\boldsymbol{\delta}}_{i,LTS}$ , a estimativa *LTS* de  $\boldsymbol{\delta}_i$  e calcularam-se os resíduos correspondentes,  $\hat{e}_{it} = y_{it} - \mathbf{Z}_{it}\hat{\boldsymbol{\delta}}_{i,LTS}$ , para  $i = 1, \dots, M$ .

### 5.3 Estimação da matriz de covariâncias $\mathbf{\Omega}$

Como se mostrou na **Secção 3.8**, foi possível encontrar uma representação explícita para o estimador *GMM* (3.8.4), a qual depende da matriz de covariâncias dos erros do modelo,  $\mathbf{\Omega}$ . Tradicionalmente, esta matriz é desconhecida e, ao usar o estimador *GMM* em modelos heterocedásticos, a estimação prévia desta matriz é feita recorrendo ao estimador de White. No entanto, por este processo, cada elemento da matriz de covariâncias é estimado com base num único resíduo, o que pode conduzir a resultados desastrosos, no caso de haver resíduos atípicos na amostra.

Na versão robusta do estimador *GMM* proposta neste trabalho, optou-se por efectuar a estimação prévia da matriz de covariâncias  $\mathbf{\Omega}$ , recorrendo a um dos métodos robustos conhecidos, neste caso recorrendo ao estimador robusto *OGK*, apresentado na **Subsecção 4.2.3**. A escolha do estimador *OGK*, em detrimento de outros estimadores robustos da matriz de covariâncias mais usados actualmente (como é o caso do *MCD*,

igualmente referido na **Subsecção 4.2.3**), teve a ver com o facto de ser necessário dispor de uma estimativa da própria matriz de covariâncias, e não de medidas de dispersão resumidas, como por exemplo o determinante da matriz estimada. Ao optar pelo estimador *OGK*, teve-se ainda em conta que cada elemento da matriz de covariâncias é estimado individualmente, o que garante que se estimam de forma robusta as covariâncias entre pares de variáveis e também a dependência entre equações do sistema. Note-se que, para estes modelos e sob a hipótese de heterocedasticidade,  $\mathbf{\Omega}$  não é uma matriz diagonal; de facto,  $\mathbf{\Omega}$  é formada por blocos diagonais com elementos distintos ao longo das diversas diagonais. Outras razões que influenciaram a escolha deste estimador foram a rapidez na obtenção dos resultados, mesmo com amostras de dimensão elevada, e ainda o facto de ser robusto e aproximadamente equivariante relativamente à localização e à escala. Em relação ao processo de cálculo das estimativas por *OGK*, têm vindo a ser desenvolvidos programas para o ambiente *R*, nomeadamente a *package rrcov*, que inclui a função *CovOgk*, que permite obter as estimativas em causa.

No entanto, a aplicação do estimador *OGK* aos *SEM* e a determinação das correspondentes estimativas, requereu um trabalho prévio de adaptação do estimador a estes modelos. Com efeito, recorde-se que  $\mathbf{\Omega}$  é composta por blocos, ou seja, é da forma

$$\mathbf{\Omega} = \begin{bmatrix} \mathbf{\Omega}_1 & \cdots & \mathbf{\Omega}_{1M} \\ \vdots & \ddots & \vdots \\ \mathbf{\Omega}_{M1} & \cdots & \mathbf{\Omega}_M \end{bmatrix}.$$

No entanto, cada bloco é diagonal, mas a diagonal é formada por elementos distintos, devido à heterocedasticidade. Logo, foi necessário dispor de várias realizações dos resíduos, para poder estimar cada um dos elementos das diversas diagonais. Para obter esses resíduos, procedeu-se à reamostragem dos resíduos iniciais. Mas o facto dos resíduos iniciais serem calculados a partir de estimadores iniciais robustos dos coeficientes, não assegura que os resíduos também o sejam. Basta notar que  $\hat{e}_{it} = y_{it} - \mathbf{Z}_{it}\hat{\boldsymbol{\delta}}_{i,LTS}$ ; logo se  $y_i$  e  $\mathbf{Z}_i$  contiverem observações atípicas, o  $\hat{e}_{it}$  pode ser arbitrariamente grande, mesmo com  $\hat{\boldsymbol{\delta}}_{i,LTS}$  robusto. Tendo em conta que a reamostragem de dados contaminados pode ampliar o efeito da contaminação, optou-se por controlar esse aspecto. Para isso, reamostraram-se apenas os resíduos correspondentes às observações que foram usadas no cálculo das estimativas *LTS* iniciais. Numa fase intermédia, estudou-se também a



possibilidade de fazer subamostragem, a qual foi abandonada porque, embora se adapte ao caso particular da homocedasticidade, sob a hipótese de heterocedasticidade, originava uma disparidade nas dimensões das amostras de resíduos, obtidos pelo método *OGK* ou pelo estimador de White, aspecto que dificultava a posterior comparação dos resultados.

Assim, conforme se viu na secção anterior, as estimativas iniciais dos parâmetros foram obtidas aplicando o estimador *LTS* a cada equação  $i$ , obtendo-se  $\hat{\delta}_{i,LTS}$ , a estimativa *LTS* para  $\delta_i$ . De seguida, como forma de controlar o efeito da contaminação na reamostragem dos resíduos, teve-se em consideração que, no processo de cálculo do estimador *LTS*, se excluem os  $(T - h_i)$  maiores resíduos absolutos, e reamostraram-se apenas os  $h_i$  resíduos usados na obtenção das estimativas de cada vector  $\hat{\delta}_{i,LTS}$ .

Ao procurar adaptar o estimador *OGK* aos *SEM*, seguiu-se inicialmente um método que posteriormente foi abandonado. O estudo inicial consistiu em usar o estimador *OGK* para obter estimativas robustas de cada um dos blocos da matriz  $\Omega$ . Nos blocos da diagonal principal,  $\Omega_i = \text{diag}[\sigma_{it}^2] = \text{diag}[E(e_{it}^2)]$ , usava-se o estimador de escala *OGK* calculado a partir de matrizes de resíduos,  $\hat{e}_{it}$ . Este procedimento permitia obter uma estimativa robusta da matriz de covariâncias destes resíduos. De seguida, com os elementos diagonais da matriz obtida, formava-se a matriz

$$\hat{\Omega}_{i,OGK} = \text{diag} \begin{bmatrix} \hat{\sigma}_{i1}^2 & \cdots & \hat{\sigma}_{iT}^2 \end{bmatrix},$$

onde  $\hat{\sigma}_{it}^2 = \widehat{V(e_{it})}$  era calculado pelo processo anteriormente descrito. Nos blocos fora da diagonal principal,  $\Omega_{ij} = \text{diag}[\sigma_{ijt}] = \text{diag}[E(e_{it}e_{jt})]$ , usava-se o estimador de localização *OGK*, partindo de matrizes formadas pelos produtos dos resíduos,  $\hat{e}_{it}\hat{e}_{jt}$ , para obter estimativas robustas  $\widehat{E(e_{it}e_{jt})}$ . De seguida, com as componentes do vector de dimensão  $T$  obtido, formava-se a matriz

$$\hat{\Omega}_{ij,OGK} = \text{diag} \begin{bmatrix} \hat{\sigma}_{ij1} & \cdots & \hat{\sigma}_{ijT} \end{bmatrix},$$

onde  $\hat{\sigma}_{ijt} = \widehat{E(e_{it}e_{jt})}$ . Usando estas estimativas dos blocos, compunha-se a estimativa global de  $\Omega$ .

No seguimento do trabalho notou-se que, pelo processo anterior não havia a garantia de que a matriz estimada  $\hat{\Omega}$  fosse sempre semidefinida positiva. Assim, o estudo evoluiu

para outro modo de adaptação do método *OGK*, seguindo o mesmo princípio, mas de forma a assegurar que a matriz estimada fosse semidefinida positiva.

Para corrigir a metodologia anteriormente investigada, e tendo em atenção a heterocedasticidade do modelo, foi necessário organizar os resíduos reamostrados por equação, formando matrizes  $\mathbf{P}_i$ , constituídas pelas realizações dos resíduos associados à equação  $i$ .

Posteriormente, juntaram-se os  $M$  blocos  $\mathbf{P}_i$  e formou-se uma matriz de resíduos alargada,  $\mathbf{P} = \begin{bmatrix} \mathbf{P}_1 & \cdots & \mathbf{P}_M \end{bmatrix}$ . Aplicando o estimador robusto de escala *OGK* à matriz  $\mathbf{P}$  obteve-se uma matriz de covariâncias estimada. Tendo em conta a estrutura da matriz  $\mathbf{\Omega}$ , relembra-se que os elementos fora das diagonais principais de cada um dos blocos são nulos. Deste modo, a matriz de resíduos reamostrados  $\mathbf{P}$  apenas foi usada para estimar os elementos não nulos de  $\mathbf{\Omega}$ . Sendo assim, cada bloco diagonal  $\hat{\mathbf{\Omega}}_{ij,OGK}$  foi formado pelos elementos homólogos do bloco homólogo da matriz estimada por *OGK* e considerando nulos os restantes elementos. Note-se que a matriz  $\hat{\mathbf{\Omega}}_{OGK}$  assim obtida é semidefinida positiva, pois tem elementos diagonais coincidentes com os da matriz estimada por *OGK*, estando provado (ver Maronna e Zamar, 2002) que esta última goza dessa propriedade.

## 5.4 Resolução do problema de minimização

Estando decidido qual o método para obter as estimativas iniciais do processo de minimização e o método de estimação robusta de  $\mathbf{\Omega}$ , foi necessário escolher qual o estimador robusto a usar para estimar os coeficientes do sistema de equações, ou seja, os parâmetros do modelo. Esta etapa corresponde à terceira e última fase do processo de estimação que se propõe.

Para prosseguir com a estimação robusta dos coeficientes das equações teve-se em conta a analogia com a estimação robusta em modelos de regressão, não esquecendo as duas diferenças fundamentais - variância dos erros não constante e existência de correlação entre os erros das diferentes equações. Como se viu na **Secção 4.3**, existem várias vias para estimar de forma robusta os coeficientes da regressão; no entanto, elas foram desenvolvidas para cenários *i.i.d.*, que não se verificam nos *SEM*. Deste modo,

analisaram-se diferentes possibilidades. Em particular, e dados os bons resultados a que têm conduzido os estimadores-MM, começou por se colocar a hipótese de os utilizar. No entanto, a vantagem desses estimadores está associada ao modo como estimam uma única constante como termo de escala; por isso, considerou-se que não eram adequados aos *SEM*, sob as hipóteses de heterocedasticidade e estimação completa em que se pretendia trabalhar. Quanto à utilização dos estimadores-*GM*, mantinham-se as mesmas dificuldades, agravadas pelo facto do ponto de ruptura ser baixo, em consequência do elevado número de parâmetros presentes no sistema. Assim, optou-se por usar um estimador de Huber, uma vez que este estimador permite incluir as estimativas de escala de cada um dos resíduos, obtidos na segunda fase do processo de estimação. Infelizmente, esta opção impediu a utilização das funções previamente programadas em *R* para o cálculo das estimativas mas, tendo em conta que a função de Huber não exige uma programação complexa, decidiu-se prosseguir com esta metodologia.

Assim, a estimação de  $\delta$  prosseguiu com a resolução de um problema de minimização que fosse uma alternativa ao processo usual de definir o estimador *GMMIC*, ou seja,

$$\min_{\delta} \left\{ [(\mathbf{I} \otimes \mathbf{X}') (\mathbf{y} - \mathbf{Z}\delta)]' \hat{\mathbf{W}} [(\mathbf{I} \otimes \mathbf{X}') (\mathbf{y} - \mathbf{Z}\delta)] \right\}, \quad (5.4.1)$$

onde  $\hat{\mathbf{W}} = \left[ T^{-1} (\mathbf{I} \otimes \mathbf{X}') \hat{\mathbf{\Omega}} (\mathbf{I} \otimes \mathbf{X}) \right]^{-1}$ .

Nesse sentido, considerou-se o problema (5.4.1) na forma

$$\min_{\delta} \left\{ (\mathbf{y} - \mathbf{Z}\delta)' (\mathbf{A}\hat{\mathbf{\Omega}}^{-1}) (\mathbf{y} - \mathbf{Z}\delta) \right\}, \quad (5.4.2)$$

obtida após operações de transposição, e considerando a matriz de dimensão  $(MT \times MT)$ , definida por

$$\mathbf{A}\hat{\mathbf{\Omega}}^{-1} = (\mathbf{I} \otimes \mathbf{X}) \left[ (\mathbf{I} \otimes \mathbf{X}') \hat{\mathbf{\Omega}} (\mathbf{I} \otimes \mathbf{X}) \right]^{-1} (\mathbf{I} \otimes \mathbf{X}'). \quad (5.4.3)$$

Para garantir a existência de uma decomposição que será necessária no seguimento, há que garantir que a matriz  $\left( \mathbf{A}\hat{\mathbf{\Omega}}^{-1} \right)^{\frac{1}{2}}$ , de dimensão  $(MT \times MT)$ , é simétrica e semi-definida positiva. Para isso, mostra-se de seguida que a matriz  $\mathbf{A}\hat{\mathbf{\Omega}}^{-1}$  é simétrica e semidefinida positiva pois, se assim for, a matriz  $\left( \mathbf{A}\hat{\mathbf{\Omega}}^{-1} \right)^{\frac{1}{2}}$  goza das mesmas propriedades.

1.  $\mathbf{A}\hat{\mathbf{\Omega}}^{-1}$  é simétrica.

*Demonstração.*  $\mathbf{A}\hat{\mathbf{\Omega}}^{-1}$  é simétrica se  $\left(\mathbf{A}\hat{\mathbf{\Omega}}^{-1}\right)' = \mathbf{A}\hat{\mathbf{\Omega}}^{-1}$ . De facto,

$$\begin{aligned}\left(\mathbf{A}\hat{\mathbf{\Omega}}^{-1}\right)' &= \left[(\mathbf{I} \otimes \mathbf{X}) \left[(\mathbf{I} \otimes \mathbf{X}') \hat{\mathbf{\Omega}} (\mathbf{I} \otimes \mathbf{X})\right]^{-1} (\mathbf{I} \otimes \mathbf{X}')\right]' \\ &= (\mathbf{I} \otimes \mathbf{X}) \left[(\mathbf{I} \otimes \mathbf{X}') \hat{\mathbf{\Omega}} (\mathbf{I} \otimes \mathbf{X})\right]^{-1} (\mathbf{I} \otimes \mathbf{X}') \\ &= \mathbf{A}\hat{\mathbf{\Omega}}^{-1}.\end{aligned}$$

□

2.  $\mathbf{A}\hat{\mathbf{\Omega}}^{-1}$  é semidefinida positiva.

*Demonstração.* Para verificar esta propriedade, utilizam-se os seguintes resultados auxiliares sobre matrizes:

P1: Se  $\mathbf{A}_{(n \times n)}$  é semidefinida positiva, então  $\mathbf{P}'\mathbf{A}\mathbf{P}$  é semidefinida positiva, para  $\mathbf{P}_{(n \times m)}$ .

P2: Se  $\mathbf{A}$  é semidefinida positiva e não singular, então  $\mathbf{A}^{-1}$  também é semidefinida positiva.

Como  $\hat{\mathbf{\Omega}}^{-1}$  é uma matriz de covariâncias, então é semidefinida positiva. Considerando que

$$\mathbf{A}\hat{\mathbf{\Omega}}^{-1} = (\mathbf{I} \otimes \mathbf{X}) \left[(\mathbf{I} \otimes \mathbf{X}') \hat{\mathbf{\Omega}} (\mathbf{I} \otimes \mathbf{X})\right]^{-1} (\mathbf{I} \otimes \mathbf{X}'),$$

fazendo  $\mathbf{P} = \mathbf{I} \otimes \mathbf{X}$  na propriedade P1, aplicando seguidamente P2 e, novamente P1, agora com  $\mathbf{P} = \mathbf{I} \otimes \mathbf{X}'$ , conclui-se que  $\mathbf{A}\hat{\mathbf{\Omega}}^{-1}$  é semidefinida positiva. □

Então, tendo em conta que se verifica

$$\mathbf{A}\hat{\mathbf{\Omega}}^{-1} = \left(\mathbf{A}\hat{\mathbf{\Omega}}^{-1}\right)^{\frac{1}{2}} \left(\mathbf{A}\hat{\mathbf{\Omega}}^{-1}\right)^{\frac{1}{2}} = \left[\left(\mathbf{A}\hat{\mathbf{\Omega}}^{-1}\right)^{\frac{1}{2}}\right]' \left(\mathbf{A}\hat{\mathbf{\Omega}}^{-1}\right)^{\frac{1}{2}},$$

e fazendo  $\mathbf{A}^* = \left[\mathbf{A}\hat{\mathbf{\Omega}}^{-1}\right]^{\frac{1}{2}}$ , pode reescrever-se o problema de minimização referido em (5.4.2) do seguinte modo:

$$\begin{aligned}
& \min_{\delta} \left\{ (\mathbf{y} - \mathbf{Z}\delta)' \left[ \left[ (\mathbf{A}\hat{\mathbf{\Omega}}^{-1}) \right]^{\frac{1}{2}} \right]' (\mathbf{A}\hat{\mathbf{\Omega}}^{-1})^{\frac{1}{2}} (\mathbf{y} - \mathbf{Z}\delta) \right\} \\
&= \min_{\delta} \{ (\mathbf{y} - \mathbf{Z}\delta)' (\mathbf{A}^*)' \mathbf{A}^* (\mathbf{y} - \mathbf{Z}\delta) \} \\
&= \min_{\delta} \{ [\mathbf{A}^* (\mathbf{y} - \mathbf{Z}\delta)]' [\mathbf{A}^* (\mathbf{y} - \mathbf{Z}\delta)] \} \\
&= \min_{\delta} \left\| \underbrace{\underbrace{\mathbf{A}^*}_{(MT \times MT)} \underbrace{(\mathbf{y} - \mathbf{Z}\delta)}_{(MT \times 1)}}_{(MT \times 1)} \right\|^2 \\
&= \min_{\delta} \sum_{j=1}^{MT} [\mathbf{A}^* (\mathbf{y} - \mathbf{Z}\delta)]_j^2,
\end{aligned}$$

onde  $[\mathbf{A}^* (\mathbf{y} - \mathbf{Z}\delta)]_j$  representa a componente de ordem  $j$  do vector  $[\mathbf{A}^* (\mathbf{y} - \mathbf{Z}\delta)]$ .

Esta última forma de enunciar o problema tem a vantagem de evidenciar que o estimador *GMM* é obtido minimizando uma função objectivo da forma

$$\min \sum_k \rho(u_k), \text{ com } \rho(u) = u^2,$$

ou seja, que o estimador *GMM* é um estimador-M. Para obter uma versão robusta do estimador, entrando com os elementos de  $\hat{\mathbf{\Omega}}_{OGK}$  previamente calculados, e sendo desconhecida a existência de métodos computacionais já programados e devidamente testados, que prevejam a forma específica de  $\mathbf{\Omega}$ , optou-se por um estimador de Huber. Deste modo, os elementos da matriz  $\hat{\mathbf{\Omega}}_{OGK}$ , desempenharam um papel fundamental para poder escalonar os resíduos na aplicação do estimador de Huber.

Assim, o problema de minimização foi resolvido considerando a função  $\rho$  como sendo a função  $\rho$  de Huber, definida no **Capítulo 3**. A versão robusta do estimador *GMM* passou a ser definida como solução de

$$\min_{\delta} \sum_{j=1}^{MT} \rho_b \left\{ \frac{[\mathbf{A}^* (\mathbf{y} - \mathbf{Z}\delta)]_j}{s_j} \right\} = \min_{\delta} \sum_{j=1}^{MT} \rho_b \left[ \frac{(\mathbf{A}^* \mathbf{e})_j}{s_j} \right],$$

onde  $s_j$  estima o desvio padrão de  $(\mathbf{A}^* \mathbf{e})_j$ . Note-se que esta função  $\rho_b$  tem argumento unidimensional, pois é aplicada a cada componente do vector  $\mathbf{A}^* \mathbf{e}$ .

Para estudar a distribuição de  $(\mathbf{A}^* \mathbf{e})_j$ , note-se que  $\mathbf{e} \sim \mathbf{N}_{MT}(\mathbf{0}, \hat{\mathbf{\Omega}})$  e por uma

propriedade da distribuição normal, então

$$\mathbf{A}^* \mathbf{e} \sim \mathbf{N}_{MT} \left( \mathbf{0}, \mathbf{A}^* \hat{\mathbf{\Omega}} (\mathbf{A}^*)' \right).$$

Deste modo, para a componente  $j$ , do vector  $\mathbf{A}^* \mathbf{e}$ , sabe-se que

$$[\mathbf{A}^* \mathbf{e}]_j \sim \mathbf{N}_1 \left( 0, \left[ \mathbf{A}^* \hat{\mathbf{\Omega}} (\mathbf{A}^*)' \right]_{jj} \right),$$

onde  $\left[ \mathbf{A}^* \hat{\mathbf{\Omega}} (\mathbf{A}^*)' \right]_{jj}$  é o elemento de ordem  $j$  da diagonal principal da matriz  $\mathbf{A}^* \hat{\mathbf{\Omega}} (\mathbf{A}^*)'$ , ou ainda que

$$\frac{[\mathbf{A}^* \mathbf{e}]_j}{\sqrt{\left[ \mathbf{A}^* \hat{\mathbf{\Omega}} (\mathbf{A}^*)' \right]_{jj}}} \sim \mathbf{N}_1 (0, 1). \quad (5.4.4)$$

Assim, sugere-se que a versão robusta do estimador *GMM* seja obtida como solução do problema

$$\min_{\delta} \sum_{j=1}^{MT} \rho_b \left\{ \frac{[\mathbf{A}^* \mathbf{e}]_j}{\sqrt{\left[ \mathbf{A}^* \hat{\mathbf{\Omega}} (\mathbf{A}^*)' \right]_{jj}}} \right\}. \quad (5.4.5)$$

Este estimador que se obtém considerando as três etapas, terminando na solução do problema (5.4.5), constitui a versão robusta do estimador *GMMIC* e que será designado por  $\hat{\delta}_{GMMOGK}$ .

Finalmente, falta decidir o valor da constante de afinação  $b$  na função de Huber. Atendeu-se ao facto do actual argumento da função  $\rho$  ser uma variável aleatória com distribuição normal reduzida, como se confirmou em (5.4.4). Deste modo, a escolha do valor da constante de afinação seguiu os critérios propostos, por exemplo em Pires e Branco (2007), para o modelo de regressão, os quais recomendam um valor da constante que assegure uma eficiência assintótica elevada, sob o modelo subjacente. Assim, fixando a eficiência assintótica de 95%, usou-se o valor  $b = 1.345$ , sob o modelo normal.

Por outro lado, no caso do modelo normal ser contaminado com uma percentagem de contaminação fixa,  $\varepsilon$ , o valor da constante de afinação pode ser calculado a partir da relação,

$$\frac{\varepsilon}{1 - \varepsilon} = \frac{2\varphi(b)}{b} - 2\Phi(-b),$$

onde  $\varphi$  e  $\Phi$  representam, respectivamente, a função densidade e a função distribuição da distribuição  $\mathbf{N}(0, 1)$ , de acordo com Pires e Branco (2007). A título de exemplo, para

o caso de uma amostra normal com contaminação 20%, obteve-se  $b = 0.861$ . Foram estes ( $b = 0.861$  e  $b = 1.345$ ) os valores usados no processo computacional descrito no capítulo seguinte, quando se trabalhou com amostras com 20% de contaminação ou com amostras sem contaminação, respectivamente.

Com o valor da constante fixado e escolhida a função  $\rho$ , a função objectivo do problema de minimização ficou perfeitamente definida. Logo, passou-se à resolução do problema, que permitiu obter as estimativas dos coeficientes do *SEM*.

No que diz respeito ao processo de optimização, necessário ao cálculo das estimativas, pode recorrer-se a uma das diversas funções pré-programadas em *R*. Neste trabalho foi utilizada a função *optim*, disponível na *package stats* do *R*.

## 5.5 Algoritmos para o cálculo de estimativas pela versão robusta do *GMM*

Nesta secção apresenta-se um algoritmo que resume o processo de cálculo das estimativas pela versão robusta do estimador *GMM*, supondo o modelo heterocedástico, o qual será designado por Algoritmo HetIC. Posteriormente, fez-se a particularização para o caso homocedástico, que se apresenta no algoritmo designado por Algoritmo HomIC. Para facilitar a implementação de alguns passos do algoritmo HetIC, apresentam-se nas Figuras 5.1 e 5.2 dois esquemas gráficos auxiliares.

### 5.5.1 Algoritmo HetIC

HetIC.1 Para cada equação  $i$  na forma  $\mathbf{y}_i = \mathbf{Z}_i \boldsymbol{\delta}_i + \mathbf{e}_i$ , obter estimativas iniciais por *LTS*,  $\hat{\boldsymbol{\delta}}_{i,LTS}$ .

HetIC.2 Para as  $h_i$  observações usadas no cálculo de  $\hat{\boldsymbol{\delta}}_{i,LTS}$ , calcular resíduos da forma  $\hat{e}_{it} = y_{it} - \mathbf{Z}_{it} \hat{\boldsymbol{\delta}}_{i,LTS}$ , para  $i = 1, \dots, M, t = 1, \dots, h_i$ .

HetIC.3 Formar uma matriz de resíduos com  $h$  linhas e  $M$  colunas, onde  $h = \min_{1 \leq i \leq M} \{h_i\}$ . Cada linha  $t$  contém o resíduo  $M$ -dimensional correspondente à observação  $t$  do sistema de  $M$  equações ( $t = 1, \dots, h$  e  $h < T$ ).

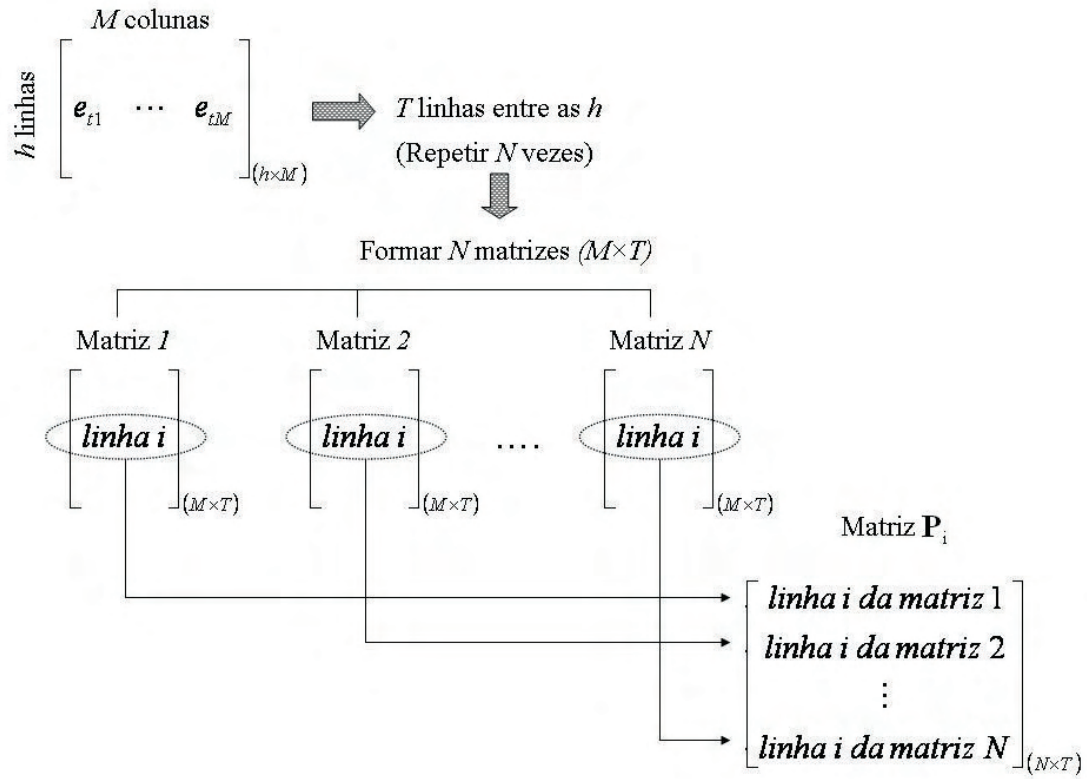


Figura 5.1: Esquema de reamostragem - passos 3 a 5 do Algoritmo HetIC.

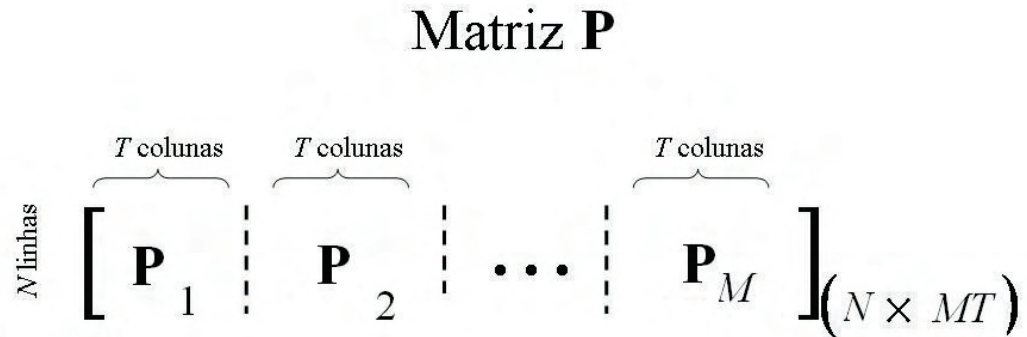


Figura 5.2: Construção da matriz  $\mathbf{P}$  - passo 6 do Algoritmo HetIC.



HetIC.4 Formar  $N$  matrizes constituídas por  $T$  resíduos  $M$ -dimensionais (com  $N$  elevado).

Para formar cada matriz, reamostrar aleatoriamente e com reposição  $T$  linhas da matriz obtida em HetIC.3.

HetIC.5 Organizar os resíduos reamostrados por equação. Para cada equação  $i$ ,  $i = 1, \dots, M$ , formar uma matriz  $\mathbf{P}_i$ , com a  $i$ -ésima linha ( $T$ -dimensional) de cada uma das  $N$  matrizes obtidas em HetIC.4. Logo,  $\mathbf{P}_i$  tem dimensão  $(N \times T)$  e é constituída por  $N$  réplicas do vector de resíduos da equação  $i$ , para as  $T$  observações.

HetIC.6 Juntar os  $M$  blocos  $\mathbf{P}_i$  ( $i = 1, \dots, M$ ), para obter uma matriz de resíduos alargada,  $\mathbf{P} = \begin{bmatrix} \mathbf{P}_1 & \dots & \mathbf{P}_M \end{bmatrix}$ , com dimensão  $(N \times MT)$ . Cada uma das  $N$  linhas de  $\mathbf{P}$  contém as  $T$  realizações de resíduos correspondentes às sucessivas equações  $i$ , para  $i = 1, \dots, M$  (cada linha é um vector de dimensão  $MT$ ). Cada coluna de  $\mathbf{P}$  contém  $N$  réplicas de cada um dos resíduos  $\hat{e}_{it}$ .

HetIC.7 Com as  $N$  réplicas de cada um dos  $MT$  resíduos, estimar a matriz de covariâncias  $(MT \times MT)$ , aplicando o estimador robusto *OGK*.

HetIC.8 Formar a matriz  $\hat{\mathbf{\Omega}}_{OGK}$  com os blocos diagonais,  $\hat{\mathbf{\Omega}}_{ij,OGK}$ , constituídos pelos elementos diagonais de índices homólogos da matriz de covariâncias estimada por *OGK* em HetIC.7 (mantendo nulos os restantes elementos).

HetIC.9 Obter a versão robusta do estimador *GMM*,  $\hat{\boldsymbol{\delta}}_{GMMOGK}$ , como solução de minimização da função

$$\sum_{j=1}^{MT} \rho_b \left\{ \frac{[\mathbf{A}^* \mathbf{e}]_j}{\sqrt{[\mathbf{A}^* \hat{\mathbf{\Omega}}_{OGK} (\mathbf{A}^*)']_{jj}}} \right\},$$

onde  $\rho_b$  é a função  $\rho$  de Huber,  $\mathbf{A}^* = [\mathbf{A} \hat{\mathbf{\Omega}}_{OGK}^{-1}]^{\frac{1}{2}}$  e  $\mathbf{A}$  é a matriz referida em (5.4.3).

O algoritmo anterior fica claramente simplificado quando se admite o caso particular do modelo homocedástico (aliás, o próprio processo de cálculo do estimador se simplifica muito, uma vez que deixa de ser necessário proceder à reamostragem dos resíduos). Apresenta-se de seguida o algoritmo correspondente a essa situação.

### 5.5.2 Algoritmo HomIC

HomIC.1 Para cada equação  $i$  na forma  $\mathbf{y}_i = \mathbf{Z}_i \boldsymbol{\delta}_i + \mathbf{e}_i$ , obter estimativas iniciais por *LTS*,  $\hat{\boldsymbol{\delta}}_{i,LTS}$ .

HomIC.2 Para as  $h_i$  observações usadas no cálculo de  $\hat{\boldsymbol{\delta}}_{i,LTS}$ , calcular resíduos da forma  $\hat{e}_{it} = y_{it} - \mathbf{Z}_{it} \hat{\boldsymbol{\delta}}_{i,LTS}$ , para  $i = 1, \dots, M, t = 1, \dots, h_i$ .

HomIC.3 Formar uma matriz de resíduos com  $h$  linhas e  $M$  colunas, onde  $h = \min_{1 \leq i \leq M} \{h_i\}$ . Cada linha contém o resíduo  $M$ -dimensional correspondente à observação  $t$  ( $t = 1, \dots, h$  e  $h < T$ ). Cada coluna contém  $h$  realizações de cada um dos  $M$  resíduos  $\hat{e}_i$ .

HomIC.4 Com as  $h$  realizações de cada um dos  $M$  resíduos, estima-se a matriz de covariâncias ( $M \times M$ ), aplicando o estimador robusto *OGK*, originando  $\hat{\boldsymbol{\Sigma}}_{OGK}$ .

HomIC.5 Com base na matriz obtida em HomIC.4, calcular a matriz  $\boldsymbol{\Omega}$  estimada por *OGK*, a partir de  $\hat{\boldsymbol{\Omega}}_{OGK} = \hat{\boldsymbol{\Sigma}}_{OGK} \otimes \mathbf{I}_T$ .

HomIC.6 Obter a versão robusta do estimador *GMM*,  $\hat{\boldsymbol{\delta}}_{GMMOGK}$ , como solução de minimização da função

$$\sum_{j=1}^{MT} \rho_b \left\{ \frac{[\mathbf{A}^* \mathbf{e}]_j}{\sqrt{[\mathbf{A}^* \hat{\boldsymbol{\Omega}}_{OGK} (\mathbf{A}^*)']_{jj}}} \right\}$$

onde  $\rho_b$  é a função  $\rho$  de Huber,  $\mathbf{A}^* = [\mathbf{A} \hat{\boldsymbol{\Omega}}_{OGK}^{-1}]^{\frac{1}{2}}$  e  $\mathbf{A}$  é a matriz referida em (5.4.3).

Desde já se chama a atenção para o facto do tempo de cálculo computacional se reduzir significativamente ao usar o algoritmo HomIC, o que naturalmente se deve à homocedasticidade, por dispensar o esquema de reamostragem.

## Capítulo 6

### Estudos de aplicação

O presente capítulo destina-se a avaliar o desempenho do estimador robusto que foi proposto no capítulo anterior, ou seja, da versão robusta do estimador *GMM*. Para esse efeito consideram-se dois exemplos de *SEM* baseados em modelos econométricos já estudados por outros autores, de modo a permitir efectuar a comparação de resultados, sempre que possível. Note-se que o objectivo do estudo estatístico que se efectua diz respeito, não à modelação econométrica, mas sim, à inferência sobre os parâmetros. Em particular, o interesse reside na estimação dos coeficientes, considerando amostras contaminadas e não contaminadas. Assim, num dos exemplos considerados foi possível usar as observações originais e comparar as estimativas agora calculadas com as anteriormente publicadas. O outro exemplo utiliza um modelo já trabalhado na literatura, mas inclui estudos de simulação das observações. Deste modo, as estimativas agora obtidas não foram calculadas usando precisamente os mesmos dados que foram usados em estudos publicados anteriormente.

O capítulo encontra-se organizado em três secções. Na primeira explicam-se os métodos usados para avaliar o desempenho dos estimadores, enquanto as duas últimas se destinam ao estudo de cada um dos exemplos. Note-se que os indicadores utilizados para a avaliação são necessariamente diferentes em cada exemplo, uma vez que no exemplo real temos observações que servem de padrão para a comparação, enquanto que no exemplo simulado é necessário recorrer a outras formas de avaliação.

## 6.1 Avaliação do desempenho da versão robusta do estimador *GMM*

Como o estimador *GMMOGK* é um estimador multivariado está disponível uma grande diversidade de técnicas, gráficas e numéricas, para avaliar o seu desempenho. A avaliação dos resultados foi feita com base no estudo da distribuição empírica dos resíduos multivariados. Os resultados finais foram avaliados de acordo com vários critérios, dois de natureza numérica e um de natureza gráfica. As opções seguidas nesse estudo foram inspiradas pelos critérios seguidos por outros autores em estudos que envolviam análise de dados em *SEMs* ou, mais geralmente, em análise de dados multivariados. De facto, perante a dificuldade de definir uma única medida do viés em estimadores multidimensionais, optou-se por seguir os critérios usados em Maronna *et al.* (2006), que sugerem que o estudo do desempenho do estimador se baseie em dois indicadores: um relativo à localização, correspondente à norma euclidiana das estimativas de localização e outro relativo à escala, definido pelo número de condição da matriz de covariâncias estimada.

Neste estudo, calcularam-se os indicadores referidos para a distribuição empírica dos resíduos. Mais especificamente, como indicador de localização usou-se a norma do vector de resíduos e como indicador de escala, utilizou-se o logaritmo do número de condição da matriz de covariâncias dos resíduos.

No exemplo com dados simulados, que se apresenta na **Secção 6.2**, usou-se replicação de amostras, supondo diferentes distribuições; portanto, houve necessidade de resumir os resultados obtidos para cada conjunto de amostras simuladas, de acordo com cada cenário considerado. Os resultados das réplicas foram sumariados do seguinte modo: em relação a medidas de localização, optou-se por calcular médias e medianas do indicador sugerido em Maronna *et al.* (2006), ou seja, médias e medianas das normas dos vectores de resíduos. Quanto à avaliação da dispersão das estimativas, ela foi efectuada através da comparação de matrizes de covariâncias dos resíduos. A comparação baseou-se no cálculo do logaritmo do número de condição da matriz, de acordo com Peña e Prieto (2001) e Maronna *et al.* (2006). Uma vez que o número de condição não é um dos indicadores mais vulgarizados, aproveita-se para relembrar a sua definição e interpretação. Chama-se número de condição de uma matriz quadrada (e representa-se

por  $k$ ) ao quociente entre o máximo valor próprio da matriz e o mínimo valor próprio, ou seja, a

$$k = \frac{\max_i (\lambda_i)}{\min_i (\lambda_i)},$$

onde  $\lambda_i$  são os valores próprios da matriz.

Encarando o número de condição como uma medida da dispersão das estimativas, baseada no viés da matriz de covariâncias estimada, a dispersão das estimativas é tanto maior quanto maior for o número de condição. Para facilitar a interpretação e a comparação de números de condição, Peña sugere que se use o logaritmo de base dez do número de condição ( $\log k$ ) como indicador. Os resultados podem interpretar-se do seguinte modo: um valor pequeno (próximo de zero) de  $\log k$  traduz pouca dispersão nas estimativas, situação que deve estar associada a um estimador robusto; enquanto um valor elevado de  $\log k$  corresponde a uma situação de maior dispersão das estimativas, devendo estar associada a um estimador pouco robusto. Aproveita-se a oportunidade para referir que o número de condição pode também ser usado noutros contextos, nomeadamente, quando há problemas de singularidade em análises de regressão, para analisar questões de colinearidade ou quando os regressores têm variâncias elevadas. Para resumir os valores obtidos nas amostras replicadas em cada cenário considerado, usaram-se de novo a média e a mediana dos valores de  $\log k$  obtidos para as matrizes de covariâncias amostrais dos resíduos, os quais foram calculados a partir de cada um dos estimadores considerados no estudo.

No exemplo que inclui dados reais não se colocou a questão da replicação de amostras, no entanto calcularam-se as mesmas medidas para os dados disponíveis.

Quanto à análise gráfica, em ambos os exemplos, a análise foi feita individualmente para cada equação. No exemplo apresentado na **Secção 6.3**, construíram-se os gráficos representando a distribuição empírica dos resíduos relativos a cada equação e para cada um dos estimadores, *GMMIC* e *GMMOGK*; calcularam-se as amplitudes das bandas definidas por  $2 \times 2.5 \times MAD$ , procedimento seguido no trabalho onde o mesmo exemplo foi analisado. Para a **Secção 6.2**, com dados simulados, a representação gráfica foi feita do mesmo modo, separadamente para cada equação do modelo e para cada distribuição considerada. No entanto, para resumir a informação relativa aos resíduos

obtidos para as diferentes réplicas, calcularam-se as medianas dos resíduos de cada equação, para cada conjunto de réplicas. De seguida, procedeu-se à representação gráfica das medianas dos resíduos de cada equação, para cada um dos estimadores considerados, bem como das bandas definidas por  $(+/-) 2.5 \times MAD$  e compararam-se as amplitudes das bandas.

A avaliação da robustez é mais evidente no primeiro exemplo, uma vez que inclui a simulação. Para cada uma das amostras simuladas, estudaram-se cinco situações com diferentes distribuições dos termos de erro. Em particular,

- (i) Modelo heterocedástico com distribuição normal multivariada, sem contaminação - NHC0.
- (ii) Modelo homocedástico com distribuição normal multivariada, sem contaminação - NHOC0.
- (iii) Modelo heterocedástico com distribuição normal multivariada, com grau de contaminação 20% - NHC20.
- (iv) Modelo homocedástico com distribuição normal multivariada, com grau de contaminação 20% - NHOC20.
- (v) Distribuições marginais t de Student com 5 graus de liberdade e matriz de covariâncias  $\Sigma$ , já usada na **Subsecção 3.10.2** -  $T_5\Sigma$ .

A análise dos resultados foi orientada tendo em conta que, quando não há contaminação um estimador robusto tem um bom desempenho quando origina estimativas próximas dos valores introduzidos para os parâmetros; e que, por outro lado, em amostras contaminadas, o estimador robusto conduz a estimativas pouco afastadas dos valores usados na simulação (ao contrário do estimador não robusto).

Nos pontos (iii) e (iv), a obtenção de amostras com 20% de contaminação foi feita substituindo 20% dos valores da variável endógena  $y_2$  por outros valores que fossem muito diferentes dos restantes. Assim, no caso homocedástico referido em (iv), a contaminação de uma amostra foi introduzida substituindo  $T \times 20\%$  valores de  $y_2$  pelo valor que se obtém multiplicando o seu desvio padrão amostral corrigido por uma

constante (que, neste caso foi 20), ou seja,  $T \times 20\%$  valores de  $y_2$  foram substituídos por  $20 \times (s_{y_2})_c$ , onde  $(s_{y_2})_c$  é o desvio padrão amostral corrigido na amostra não contaminada. No caso heterocedástico referido em (iii), foi necessário introduzir a contaminação doutro modo, uma vez que a diferentes observações podem corresponder diferentes variâncias dos termos de erro. Ou seja, pelo facto da variância dos erros variar com o índice  $t$  ( $t = 1, \dots, T$ ), não seria adequado usar a estimativa constante dada por  $(s_{y_2})_c$ . Por isso, multiplicou-se a constante já usada no caso homocedástico pelo máximo das normas das linhas de  $\mathbb{X}_2$  - note-se que a matriz  $\mathbb{X}_2$  contém os valores das variáveis exógenas que são regressores na segunda equação, onde cada linha corresponde a uma observação; assim, tomou-se o valor máximo e multiplicou-se pelo desvio padrão de  $e_2$  no caso homocedástico, isto é, substituíram-se 20% dos valores de  $y_2$  por  $20 \times \sqrt{0.7} \max_t (\|\mathbb{X}_2[t, \cdot]\|)$ .

Embora se tenham calculado as médias e as medianas de ambos os indicadores, de localização e de escala, apenas se apresentam os valores das médias por serem muitos semelhantes aos das medianas.

Em relação ao exemplo da **Secção 6.3**, que contém dados reais, a avaliação da robustez foi reduzida à identificação de eventuais observações atípicas e à comparação das estimativas obtidas com o estimador *GMM* robusto e com o estimador *GMMIC* tradicional (não robusto). Uma vez que os resultados encontrados foram comparados com os publicados no artigo de Maronna e Yohai (1997), os métodos gráficos utilizados foram semelhantes aos usados nesse trabalho.

De acordo com o exposto, os resultados foram analisados seguindo duas metodologias, uma de natureza numérica - normas dos resíduos e logaritmo dos número de condição das matrizes de covariâncias dos resíduos - e outra, de natureza gráfica - representação da distribuição empírica dos resíduos relativos a cada equação e a cada um dos métodos de estimação. Note-se, no entanto, que os indicadores numéricos permitem avaliar a qualidade das estimativas de todo o sistema, enquanto que o método gráfico permite avaliar a qualidade das estimativas de cada equação separadamente.

## 6.2 Exemplo proposto por Judge *et al.* (1988) - dados simulados

O primeiro exemplo consiste na análise do *SEM* proposto por Judge *et al.* (1988), já apresentado na **Subsecção 3.10.2**. Embora o tratamento anterior já tenha envolvido simulação de dados, as condições distribucionais agora simuladas foram diferentes das adoptadas anteriormente, tendo em conta os objectivos do presente capítulo. Agora interessa evidenciar as propriedades de robustez do estimador proposto no **Capítulo 5**. Este estudo inclui três componentes que importa destacar: a primeira diz respeito ao processo de simulação das amostras de acordo com o *SEM*; a segunda é relativa à possibilidade de contaminação, às dificuldades que decorrem da heterocedasticidade e à vantagem de poder comparar as estimativas obtidas com os parâmetros usados na simulação; finalmente, a terceira é referente à replicação das simulações efectuadas, que não é trivial, em situações de heterocedasticidade.

Os resultados finais foram avaliados de acordo com diversos critérios. De acordo com o exposto na secção anterior, como indicadores de localização, calcularam-se as médias e as medianas das normas dos resíduos. Separadamente, para cada equação do modelo, procedeu-se à representação gráfica das medianas dos resíduos. Note-se que este tipo de gráfico foi usado por Maronna e Yohai (1997) mas esse estudo não envolvia replicação. Relativamente à escala, calcularam-se também médias e medianas dos logaritmos dos números de condição das matrizes de covariâncias dos resíduos.

Para cada situação de contaminação introduzida, apresentam-se, comparam-se e comentam-se os resultados, em termos de vantagens mais ou menos evidentes da estimação robusta.

*Exemplo 6.2.1.* Retome-se o modelo econométrico já tratado na **Subsecção 3.10.2**, que pode ser apresentado na seguinte forma:

$$\left\{ \begin{array}{llll} -\mathbf{Y}_1 & +\mathbf{Y}_2\gamma_{21}+\mathbf{Y}_3\gamma_{31}+\mathbf{X}_1\beta_{11} & & +\mathbf{e}_1 = 0 \\ \mathbf{Y}_1\gamma_{12}-\mathbf{Y}_2 & & +\mathbf{X}_1\beta_{12}+\mathbf{X}_2\beta_{22}+\mathbf{X}_3\beta_{32}+\mathbf{X}_4\beta_{42} & +\mathbf{e}_2 = 0 \\ & \mathbf{Y}_2\gamma_{23}-\mathbf{Y}_3 & +\mathbf{X}_1\beta_{13}+\mathbf{X}_2\beta_{23} & +\mathbf{X}_5\beta_{53}+\mathbf{e}_3 = 0 \end{array} \right. .$$

Para estudar este exemplo, tendo em vista a análise da robustez do estimador proposto no trabalho, começou-se por fazer a simulação de amostras de acordo com o modelo,



admitindo cenários diferentes dos considerados anteriormente, uma vez que visam um objectivo diferente. O procedimento seguido foi o explicado na **Subsecção 3.10.2**, que se passa a relembrar: foram geradas amostras constituídas por 120 observações do modelo, usando as matrizes dos coeficientes estruturais  $\Gamma$  e  $B$  e os valores da variáveis exógenas,  $X$ , de acordo com o definido em Judge *et al.* (1988). Note-se que as observações de  $Y$  decorreram dos valores simulados para os erros, admitindo cinco situações distribucionais distintas.

Outra parte deste trabalho, incidiu na investigação sobre a forma de introduzir a contaminação nas amostras. Recorde-se que nos *SEM* as variáveis endógenas funcionam como variáveis explicativas em algumas equações, o que dissimula o efeito da contaminação, dificultando assim a avaliação do impacto provocado pela mesma nos resultados. No sentido de procurar observações atípicas com grande influência nas estimativas não robustas, mas que devam afectar muito menos as estimativas obtidas por métodos robustos, foram efectuadas diversas experiências envolvendo a geração de amostras contaminadas, seguidas do cálculo das estimativas dos parâmetros. De entre os muitos ensaios efectuados na geração de amostras contaminadas, e que não são apresentados nesta tese, podem referir-se, por exemplo, os seguintes: contaminação de uma percentagem de observações em todas as variáveis endógenas; contaminação de uma percentagem de observações apenas para algumas variáveis endógenas; contaminação substituindo uma percentagem de observações de uma só variável endógena por valores aberrantes, fazendo depois variar a variável endógena.

Por outro lado, houve necessidade de ultrapassar vários tipos de dificuldades inerentes à geração de amostras sob heterocedasticidade. As dificuldades devidas ao facto de se trabalhar com modelos multivariados com dependência entre as equações e ainda as que decorrem da introdução de contaminação em observações que apresentam uma grande variabilidade, pelo facto de terem sido simuladas de acordo com distribuições heterocedásticas.

Por fim, procedeu-se à replicação das simulações efectuadas. Apresentam-se apenas algumas das situações estudadas durante o trabalho, que pareceram conduzir a resultados mais elucidativos e relevantes para as conclusões.

## 1. Geração dos termos de erro.

Como o sistema é formado por três equações, para gerar amostras de dimensão 120, geraram-se 120 valores de erros tridimensionais do modelo, de acordo com as seguintes condições amostrais:

NHC0 Distribuição  $N_3(\mathbf{0}, \Sigma_{H,t})$ — caso heterocedástico sem contaminação.

Inicialmente pensou-se introduzir heterocedasticidade fazendo variar as variâncias dos erros de todas as equações; no entanto atendendo às características da matriz de covariâncias do modelo (apresenta uma grande disparidade entre as variâncias), e a eventuais problemas de mau condicionamento, optou-se por transformar os dados (os valores de  $x_j$  e de  $y_i$ ) de acordo com a expressão que se apresenta de seguida e que é recomendada em situações análogas,

$$\frac{x_{jt} - \bar{x}_j}{\sqrt{T-1}s_c(x_j)}.$$

Uma vez que a transformação não conduziu a uma melhoria dos resultados, decidiu-se continuar com os dados em formato original.

A heterocedasticidade foi introduzida do seguinte modo: como a variável  $x_1$  era a única variável exógena que era regressor da primeira equação e que tomava sempre valores unitários, não se alterou a variância dos erros nesse caso; para as restantes equações, utilizou-se

$$V(e_{it}) = V(\mathbf{e}_i) \|\mathbb{X}_i[t,]\|^2,$$

onde  $\mathbb{X}_i$  é a matriz formada pelas colunas de  $\mathbf{X}$  que correspondem às variáveis exógenas que são regressores na equação  $i$ , para  $i = 2, 3$ , e  $\mathbb{X}_i[t,]$  representa a linha  $t$  de  $\mathbb{X}_i$ . A ideia foi adaptar ao caso multidimensional um procedimento frequentemente observado no caso unidimensional para provocar heterocedasticidade e que consiste em tomar  $V(\varepsilon_i) = V(\varepsilon) x_i^2$ .

Para este exemplo, vem então que

$$\begin{aligned} \sigma_{2,t}^2 &= V(e_{2t}) \\ &= V(\mathbf{e}_2) \|\mathbb{X}_2[t,]\|^2 \\ &= 0.7 \|\mathbb{X}_2[t,]\|^2, \end{aligned}$$

onde  $\mathbb{X}_2$  uma matriz por blocos do tipo  $\mathbb{X}_2 = \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \mathbf{X}_3 & \mathbf{X}_4 \end{bmatrix}_{(120 \times 4)}$ .

Em relação à terceira equação, utilizou-se

$$\begin{aligned}\sigma_{3,t}^2 &= V(e_{3t}) \\ &= V(\mathbf{e}_3) \|\mathbb{X}_3[t,]\|^2 \\ &= 15.8 \|\mathbb{X}_3[t,]\|^2,\end{aligned}$$

com  $\mathbb{X}_3 = \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \mathbf{X}_5 \end{bmatrix}_{(120 \times 3)}$ .

Concretizando, veja-se por exemplo o caso em que  $i = 3$  e  $t = 1$ : tem-se

$$V(e_{31}) = 15.8 \|\mathbb{X}_3[1,]\|^2 = 15.8 (x_{11}^2 + x_{21}^2 + x_{51}^2).$$

Os valores das covariâncias foram estabelecidos de forma a que os respectivos coeficientes de correlação se mantivessem aproximadamente iguais aos homólogos da matriz  $\mathbf{CO}$ , que é a matriz de correlação correspondente a  $\Sigma = [\sigma_{ij}]$ , ou seja,

$$\mathbf{CO} = [\rho_{ij}] = \begin{bmatrix} 1 & 0.727 & -0.95 \\ 0.727 & 1 & -0.583 \\ -0.95 & -0.583 & 1 \end{bmatrix},$$

onde  $\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$ . Assim, obteve-se

$$\begin{aligned}\rho_{12,t} &= \frac{\sigma_{12,t}}{\sigma_{1,t} \sigma_{2,t}} \\ \Leftrightarrow 0.727 &= \frac{\sigma_{12,t}}{\sqrt{227.6} \sqrt{0.7 \|\mathbb{X}_2[t,]\|^2}} \\ \Leftrightarrow \sigma_{12,t} &= 0.727 \sqrt{227.6} \sqrt{0.7 \|\mathbb{X}_2[t,]\|^2}.\end{aligned}$$

Procedendo da mesma forma, obtiveram-se as seguintes expressões para  $\rho_{13,t}$  e  $\rho_{23,t}$ , respectivamente,  $\sigma_{13,t} = -0.95 \sqrt{227.6} \sqrt{15.8 \|\mathbb{X}_3[t,]\|^2}$  e  $\sigma_{23,t} = -0.583 \sqrt{0.7 \|\mathbb{X}_2[t,]\|^2} \sqrt{15.8 \|\mathbb{X}_3[t,]\|^2}$ .

Uma vez que a matriz de covariâncias é simétrica, a matriz usada na geração dos erros foi

$$\Sigma_{H,t} = \begin{bmatrix} 227.6 & \sigma_{12,t} & \sigma_{13,t} \\ \sigma_{12,t} & \sigma_{2,t}^2 & \sigma_{23,t} \\ \sigma_{13,t} & \sigma_{23,t} & \sigma_{3,t}^2 \end{bmatrix}, \forall t = 1, \dots, 120.$$

NHOC0 Distribuição  $N_3(\mathbf{0}, \Sigma)$ — caso homocedástico sem contaminação.

Nestas condições não houve necessidade de introduzir cálculos adicionais.

NHC20 Distribuição  $N_3(\mathbf{0}, \Sigma_{H,t})$ — caso heterocedástico, com grau de contaminação 20%.

A ideia que presidiu a esta situação foi a de introduzir valores de  $y_2$  que fossem muito díspares dos restantes. Para conseguir obter essas observações atípicas, substituíram-se 20% dos valores de  $y_2$  por  $20 \times \sqrt{0.7} \max_t (\|\mathbb{X}_2[t, \cdot]\|)$ . Como neste caso a variância varia com  $t$ , calcularam-se as normas dos 120 vectores definidos pelas linhas de  $\mathbb{X}_2$  (matriz que contém os valores das variáveis exógenas que são regressores na segunda equação), tomou-se o máximo das normas, o qual foi multiplicado pelo desvio padrão de  $e_2$  correspondente ao caso homocedástico e por uma constante. Neste caso, a constante usada foi 20 e o desvio padrão referido é  $\sqrt{0.7}$ , calculado com base na matriz  $\Sigma$ , definida em (3.10.4).

NHOC20 Distribuição  $N_3(\mathbf{0}, \Sigma)$ — caso homocedástico, com grau de contaminação 20%.

A contaminação foi introduzida substituindo 20% valores de  $y_2$  pelo valor que se obteve por  $20 \times (s_{y_2})_c$ , onde  $(s_{y_2})_c$  representa o desvio padrão amostral corrigido de  $y_2$ .

$T_5 \Sigma$  Distribuições marginais  $t$  de Student com 5 graus de liberdade e matriz de covariâncias  $\Sigma$ .

## 2. Geração das amostras.

A partir dos valores gerados dos erros, simularam-se os valores das variáveis endógenas, permitindo a obtenção da totalidade da amostra. Uma vez gerada uma amostra e admitindo o cenário geral do modelo heterocedástico, importa notar as dimensões dos parâmetros em causa – no exemplo considerado, para um sistema de 3 equações, envolvendo 12 coeficientes a estimar e, para uma amostra de 120 observações, o processo envolve matrizes de covariâncias com dimensão  $(360 \times 360)$ .

## 3. Replicação das amostras.

O processo foi replicado 500 vezes, a partir de outras tantas gerações de 120 erros tridimensionais, com base em cada uma das cinco condições distribucionais anteriormente referidas, conduzindo ao tratamento de um total de 2500 amostras.

#### 4. Obtenção das estimativas dos parâmetros.

Calcularam-se estimativas dos parâmetros estruturais, aplicando os métodos *3SLS*, *GMMIC* e *GMMOGK*, a partir de cada uma das 2500 amostras simuladas e obtiveram-se os resíduos tridimensionais correspondentes a cada caso.

O cálculo das estimativas por *3SLS* e por *GMMIC* não será aqui detalhado, uma vez que já foi explicado na **Subsecção 3.10.2**. Explica-se de seguida com mais detalhe o processo de obtenção das estimativas pela versão robusta *GMMOGK*, proposta no capítulo anterior. Para cada uma das amostras de dimensão 120, estimaram-se os parâmetros estruturais, por *LTS* aplicado a cada equação, e calcularam-se os  $h$  valores dos resíduos tridimensionais correspondentes, com  $h$  definido em HetIC.3, determinados por

$$\hat{e}_{it} = y_{it} - \mathbf{Z}_{it}\hat{\boldsymbol{\delta}}_{i,LTS}, i = 1, 2, 3, t = 1, \dots, h.$$

Como se referiu na **Secção 5.3**, para poder aplicar o estimador *OGK* supondo heterocedasticidade, foi necessário fazer reamostragem sobre os resíduos anteriores. Assim, obtiveram-se 1000 reamostras aleatórias de dimensão 120 desses resíduos. Para agrupar convenientemente os resíduos reamostrados por equação, formaram-se matrizes  $\mathbf{P}_i$  ( $1000 \times 120$ ),  $i = 1, 2, 3$ , com os resíduos reamostrados correspondentes à equação  $i$ ,

$$\mathbf{P}_i = \begin{bmatrix} (\hat{e}_{i,1})_1 & \dots & (\hat{e}_{i,120})_1 \\ \vdots & & \vdots \\ (\hat{e}_{i,1})_{1000} & \dots & (\hat{e}_{i,120})_{1000} \end{bmatrix}.$$

Juntando os blocos, formou-se a matriz  $\mathbf{P}$  que se apresenta de seguida,

$$\begin{aligned} \mathbf{P} &= \begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_2 & \mathbf{P}_3 \end{bmatrix} \\ &= \begin{bmatrix} (\hat{e}_{1,1})_1 & \dots & (\hat{e}_{1,120})_1 & \dots & (\hat{e}_{3,1})_1 & \dots & (\hat{e}_{3,120})_1 \\ \vdots & & \vdots & & \vdots & & \vdots \\ (\hat{e}_{1,1})_{1000} & \dots & (\hat{e}_{1,120})_{1000} & \dots & (\hat{e}_{3,1})_{1000} & \dots & (\hat{e}_{3,120})_{1000} \end{bmatrix}. \end{aligned}$$

$\mathbf{P}$  é uma matriz por blocos, de dimensão  $(1000 \times 360)$ , contendo 1000 realizações de resíduos  $\hat{\mathbf{e}}_{it}$ , cada um dos quais com dimensão 360. À matriz  $\mathbf{P}$  agora formada, aplicou-se o estimador robusto de escala *OGK*, originando uma matriz de covariâncias estimada de dimensão  $(360 \times 360)$ . Recorde-se que, neste caso, quando o *SEM* é representado na forma  $\mathbf{y} = \mathbf{Z}\boldsymbol{\delta} + \mathbf{e}$ , referida em (3.4.2), e quando se verificam as hipóteses especificadas no modelo, a matriz de covariâncias dos erros é do tipo

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Omega}_1 & \cdots & \boldsymbol{\Omega}_{13} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Omega}_{31} & \cdots & \boldsymbol{\Omega}_3 \end{bmatrix},$$

onde  $\boldsymbol{\Omega}_{ij}$ ,  $i, j = 1, 2, 3$ , são blocos diagonais com elementos diferentes. Atendendo a essa estrutura, considerou-se uma partição semelhante para a matriz de covariâncias estimada, ou seja, considerou-se  $\hat{\boldsymbol{\Omega}}$  formada por nove blocos de dimensão  $(120 \times 120)$ . Para obter as estimativas dos blocos  $\hat{\boldsymbol{\Omega}}_{ij,OGK}$ , usaram-se os elementos homólogos das diagonais dos blocos correspondentes da matriz estimada por *OGK* e consideraram-se nulos os restantes elementos. Juntando ordenadamente todos os blocos estimados, obteve-se  $\hat{\boldsymbol{\Omega}}_{OGK}$ , uma estimativa robusta global de  $\boldsymbol{\Omega}$ . Recorde-se que, para obter  $\hat{\boldsymbol{\delta}}_{GMMOGK}$ , que estima  $\boldsymbol{\delta}$  no modelo (3.4.2), ainda falta resolver o problema de minimização definido pela expressão (5.4.5). Como nesse problema, a função objectivo depende de  $\hat{\boldsymbol{\Omega}}$  através da matriz  $\mathbf{A}^*$ , por

$$\mathbf{A}^* = \left\{ (\mathbf{I} \otimes \mathbf{X}) \left[ (\mathbf{I} \otimes \mathbf{X}') \hat{\boldsymbol{\Omega}} (\mathbf{I} \otimes \mathbf{X}) \right]^{-1} (\mathbf{I} \otimes \mathbf{X}') \right\}^{\frac{1}{2}}$$

usou-se a estimativa anteriormente obtida, designada por  $\hat{\boldsymbol{\Omega}}_{OGK}$  para concretizar  $\hat{\boldsymbol{\Omega}}$ .

Para obter a solução do problema de minimização, recorreu-se à função *optim* do programa *R*, tendo, no entanto, havido necessidade de resolver algumas questões relacionadas com o processo de convergência e com a escolha da solução inicial. Para resolver dificuldades de falta de convergência, procedeu-se a um ajustamento do número máximo de iterações configurado no programa. Como solução inicial usou-se a estimativa obtida por *LTS*, a qual conduziu a boas soluções iniciais e facilitou o processo de convergência.

## 5. Cálculo dos resíduos.

A análise dos resultados foi estabelecida com base nos resíduos, os quais foram obtidos para cada método, através das expressões:

$$\hat{\mathbf{e}}_{3SLS} = \mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\delta}}_{3SLS}, \hat{\mathbf{e}}_{GMMIC} = \mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\delta}}_{GMMIC} \text{ e } \hat{\mathbf{e}}_{GMMOGK} = \mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\delta}}_{GMMOGK}.$$

Note-se que cada um dos resíduos anteriores é um vector de dimensão 360, correspondente à notação  $\hat{\mathbf{e}} = \text{vec}\hat{\mathbf{E}}$ . Por isso, para cada método de estimação considerado, obtiveram-se de seguida os 120 resíduos tridimensionais, com os quais se formaram matrizes de resíduos, de dimensão  $(120 \times 3)$  :  $\hat{\mathbf{E}}_{3SLS}$ ,  $\hat{\mathbf{E}}_{GMMIC}$  e  $\hat{\mathbf{E}}_{GMMOGK}$ . Foram estas matrizes que serviram de base à análise de resíduos efectuada no ponto seguinte.

## 6. Análise dos resultados.

Os resultados finais foram avaliados de acordo com três indicadores - calcularam-se os dois indicadores numéricos, anteriormente referidos, e fez-se a análise gráfica das medianas dos resíduos para cada uma das equação do modelo.

Relembra-se que, como indicador de localização, se usou a média das normas dos 120 resíduos tridimensionais e, como indicador de escala, se utilizou o logaritmo do número de condição da matriz de covariâncias dos resíduos. Para obter esses valores, calculou-se a média das 120 normas dos resíduos tridimensionais e o logaritmo de base 10 do número de condição da matriz de covariâncias da amostra dos mesmos 120 resíduos. Assim, para cada amostra, calcularam-se a média das normas dos resíduos e o logaritmo do número de condição da matriz de covariâncias dos resíduos.

O procedimento foi repetido para cada uma das situações amostrais consideradas, tendo sido feitas 500 réplicas em cada caso. Para resumir essa informação, obtiveram-se os valores da média e da mediana das 500 médias das normas dos resíduos ( $\mathbf{nvr}$ ), e dos 500 logaritmos dos números de condição ( $\log(k)$ ) das matrizes de covariâncias dos resíduos. Efectuaram-se cálculos semelhantes para cada um dos cinco cenários distribucionais anteriormente referidos.

Nas Tabelas 6.1 e 6.2 resumem-se os resultados obtidos com cada método, respectivamente, para os casos heterocedástico e homocedástico, sem contaminação e com 20%

caso heterocedástico				
	sem contaminação		com 20% de contaminação	
	nvr	$\log(k)$	nvr	$\log(k)$
<i>3SLS</i>	1.5336	5.0787	1.5176	4.8689
<i>GMMIC</i>	1.5342	5.0794	1.5159	4.8366
<i>GMMOGK (HetIC)</i>	1.5698	4.1630	1.5393	4.4959

Tabela 6.1: Médias dos indicadores - caso heterocedástico - valores do GMMOGK obtidos com o algoritmo HetIC.

caso homocedástico				
	sem contaminação		com 20% de contaminação	
	nvr	$\log(k)$	nvr	$\log(k)$
<i>3SLS</i>	1.5910	0.1783	1.5965	6.1585
<i>GMMIC</i>	1.5909	0.1784	1.5826	5.9313
<i>GMMOGK (HetIC)</i>	1.5919	0.2865	1.6803	4.4963

Tabela 6.2: Médias dos indicadores - caso homocedástico - valores do GMMOGK obtidos com o algoritmo HetIC.

de contaminação.

Observando a Tabela 6.1 e a Tabela 6.2, começa por se notar que a média das normas dos resíduos é da mesma ordem de grandeza, para todos os estimadores e para todos os cenários (ver colunas **nvr**). Quanto ao logaritmo do número de condição, repare-se que, no caso heterocedástico, os menores valores foram os obtidos com a versão robusta do estimador *GMM* (ver colunas  $\log(k)$ ). Observou-se que houve redução do logaritmo do número de condição em relação ao estimador *GMM* tradicional, tanto sem contaminação como com a contaminação de 20%. No entanto a redução maior observou-se no caso de não haver contaminação. Estes resultados mostram que, usando o número de condição como indicador, a versão robusta *GMMOGK* é superior na estimação de modelos heterocedásticos, mesmo quando não existe contaminação.

Resumindo, o estimador *GMMOGK* mostra um desempenho semelhante ao *GMMIC* no que diz respeito à localização e melhor desempenho no que diz respeito à dispersão, quer exista ou não contaminação.

A análise da Tabela 6.2 não conduz às mesmas conclusões para o caso homocedástico. De facto, do ponto de vista da localização, os resultados mantêm-se da mesma ordem de grandeza, enquanto que na dispersão há um decréscimo do logaritmo do



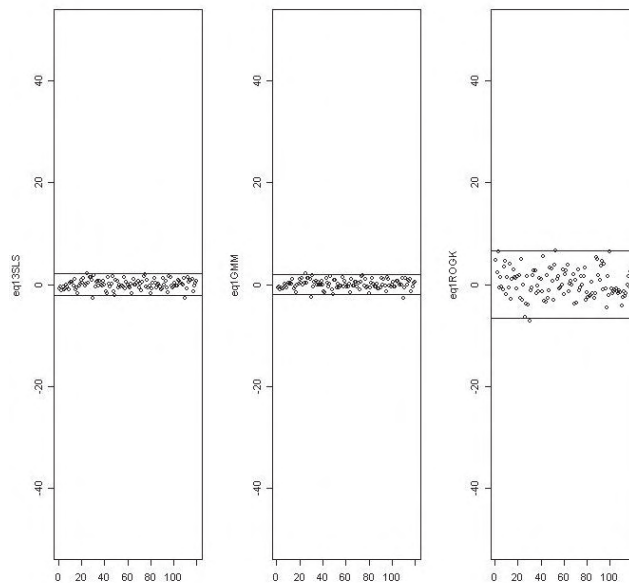


Figura 6.1: Resíduos da equação 1 obtidos com HetIC - caso heterocedástico sem contaminação.

número de condição, mas apenas no caso em que houve contaminação. A análise efectuada sobre os resultados destes dois indicadores conduz à conclusão de que a versão robusta *GMMOGK*, desenvolvida para o modelo heterocedástico, apresenta um melhor desempenho precisamente nesse caso, não trazendo tantas vantagens no caso particular homocedástico.

Para efectuar uma análise gráfica, representaram-se separadamente os resultados correspondentes a cada uma das equações. Para cada equação, apresentaram-se os 120 valores das medianas correspondentes às 120 amostras de 500 valores de resíduos de cada equação. Foram também representadas duas bandas definidas pelos limites por  $l_1 = -2.5 \times MAD$  e  $l_2 = +2.5 \times MAD$ . Este tipo de gráfico foi usado por Maronna e Yohai (1997) para representar e analisar os resultados, por equação, neste tipo de modelos. Inicialmente analisaram-se os resultados obtidos a partir das amostras correspondentes aos cenários NHC0 e NHC20. Os gráficos das medianas por equação, para o caso heterocedástico sem contaminação, apresentam-se nas Figuras 6.1 a 6.3.

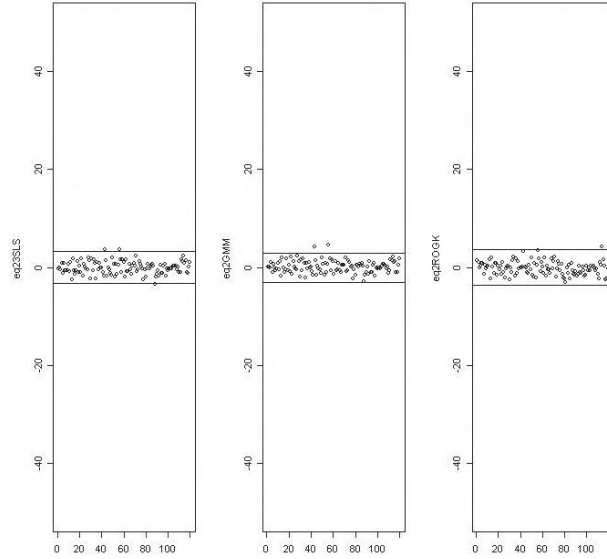


Figura 6.2: Resíduos da equação 2 obtidos com HetIC - caso heterocedástico sem contaminação.

Os gráficos das medianas por equação, para o caso heterocedástico com 20% de contaminação, apresentam-se nas Figuras 6.4 a 6.6.

Nas Figuras 6.7 a 6.9 e 6.10 a 6.12 encontra-se representado o caso homocedástico, respectivamente, sem contaminação e com 20% de contaminação, para as três equações do modelo.

Ao comparar as bandas, por equação, nota-se que não há uma superioridade notória do *GMMOGK* para todas as equações, o que leva a crer que a estrutura do sistema e de cada equação em particular é influente do desempenho do estimador. Chama-se a atenção para o facto desta conclusão ser válida para os restantes estimadores, pelo que não será certamente uma desvantagem específica do *GMMOGK*.

Os resultados numéricos encontram-se organizados nas Tabelas 6.3 e 6.4 e correspondem às amplitudes das bandas obtidas com cada método, para cada equação, respectivamente, para os casos heterocedástico e homocedástico, sem contaminação e com 20% de contaminação.

Uma vez que o modelo homocedástico pode ser visto como um caso particular do

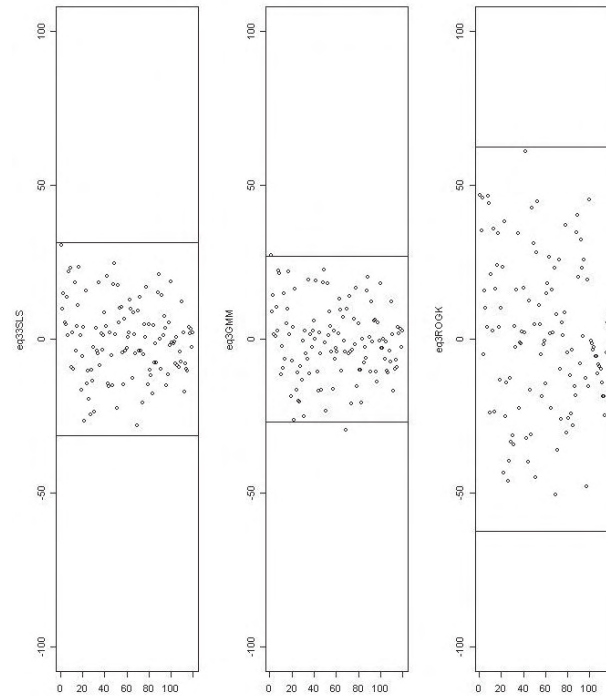


Figura 6.3: Resíduos da equação 3 obtidos com HetIC - caso heterocedástico sem contaminação.

caso heterocedástico						
	sem contaminação			com 20% de contaminação		
	equação 1	equação 2	equação 3	equação 1	equação 2	equação 3
<i>3SLS</i>	4.4493	6.6620	62.9997	284.2105	723.6842	91.0714
<i>GMMIC</i>	3.9126	5.9930	53.7879	289.4737	710.5263	89.2857
<i>GMMOGK (HetIC)</i>	13.4052	7.1255	124.8274	231.5789	657.8947	148.2143

Tabela 6.3: Largura das bandas - caso heterocedástico - valores do GMMOGK obtidos com o algoritmo HetIC.

caso homocedástico						
	sem contaminação			com 20% de contaminação		
	equação 1	equação 2	equação 3	equação 1	equação 2	equação 3
<i>3SLS</i>	6.5204	0.2493	1.2656	290.4393	730.2987	90.2108
<i>GMMIC</i>	6.3214	0.2619	1.2235	292.6451	738.0805	84.9128
<i>GMMOGK (HetIC)</i>	15.7431	0.2684	1.7628	257.8732	694.5912	149.2494

Tabela 6.4: Largura das bandas - caso homocedástico - valores do GMMOGK obtidos com o algoritmo HetIC.

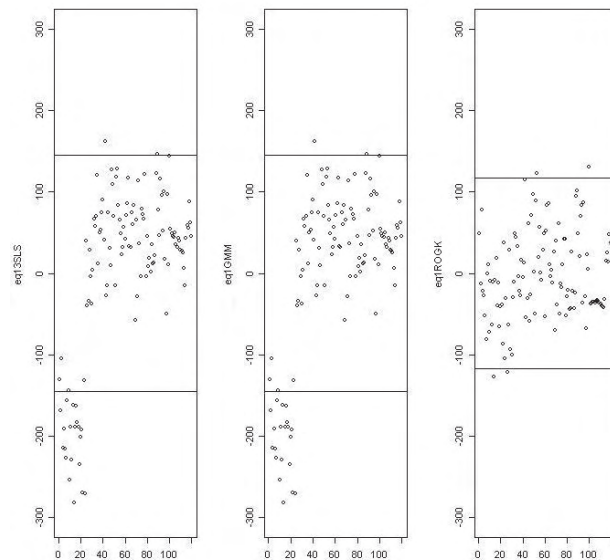


Figura 6.4: Resíduos da equação 1 obtidos com HetIC - caso heterocedástico com 20% de contaminação.

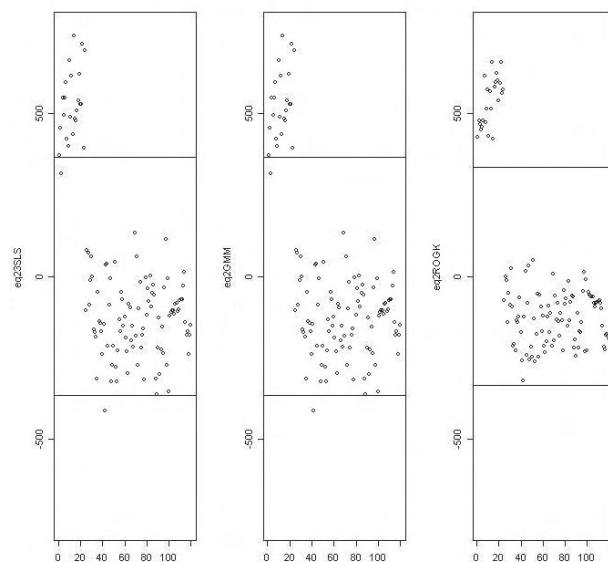


Figura 6.5: Resíduos da equação 2 obtidos com HetIC - caso heterocedástico com 20% de contaminação.

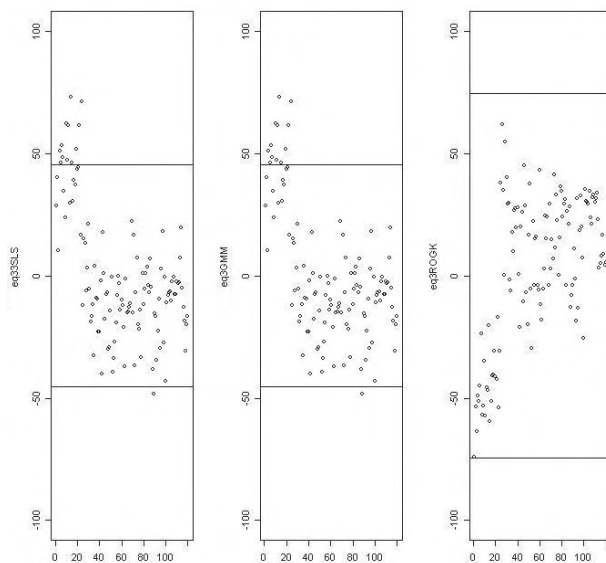


Figura 6.6: Resíduos da equação 3 obtidos com HetIC - caso heterocedástico com 20% de contaminação.

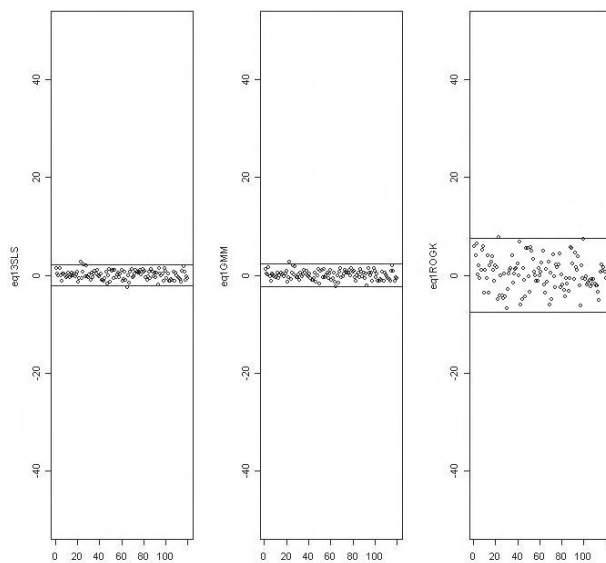


Figura 6.7: Resíduos da equação 1 obtidos por HetIC - caso homocedástico sem contaminação.

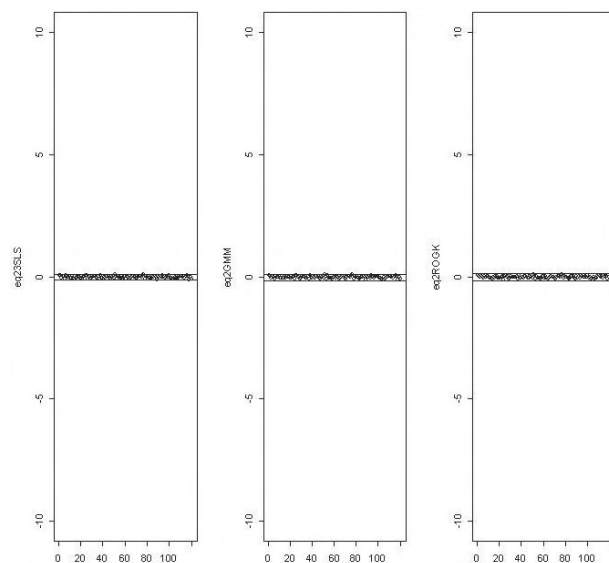


Figura 6.8: Resíduos da equação 2 obtidos por HetIC - caso homocedástico sem contaminação.

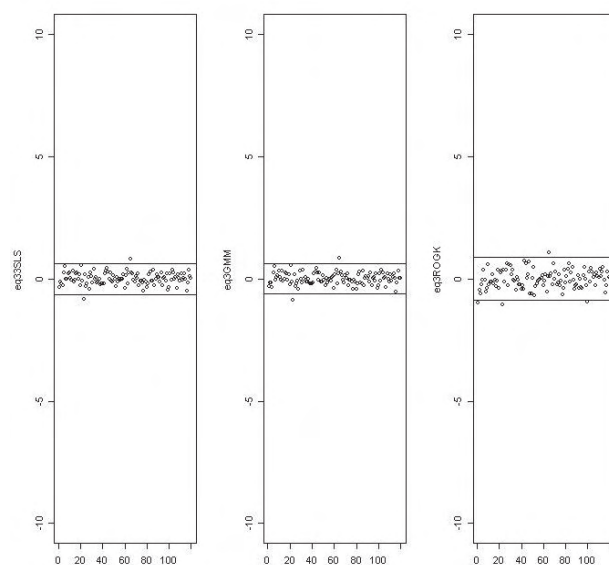


Figura 6.9: Resíduos da equação 3 obtidos por HetIC - caso homocedástico sem contaminação.

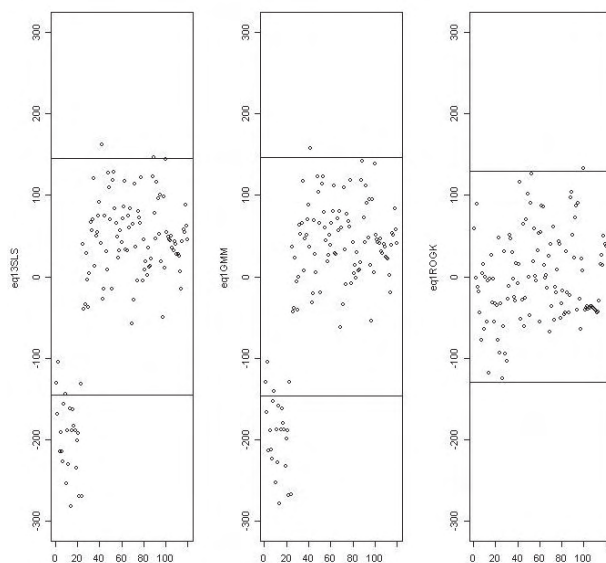


Figura 6.10: Resíduos da equação 1 obtidos por HetIC - caso homocedástico com 20% de contaminação.

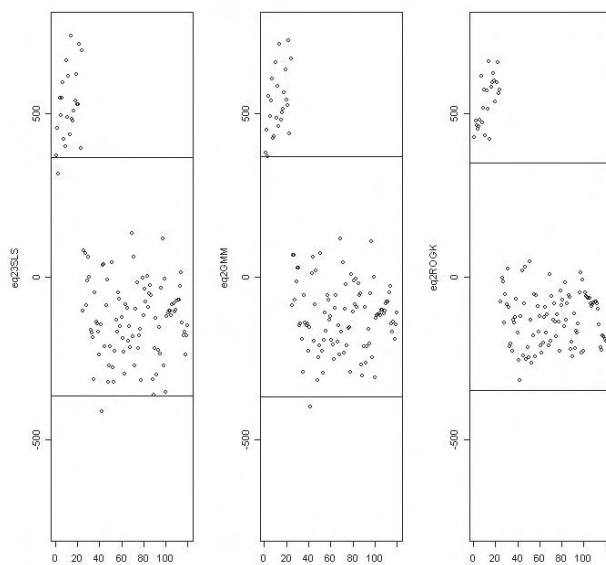


Figura 6.11: Resíduos da equação 2 obtidos por HetIC - caso homocedástico com 20% de contaminação.

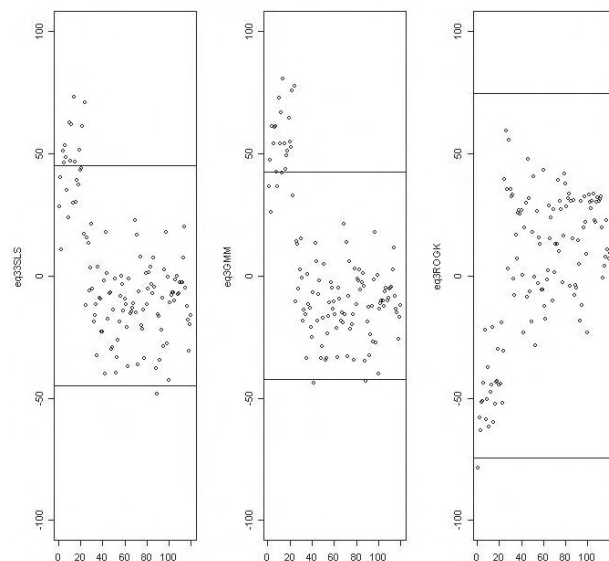


Figura 6.12: Resíduos da equação 3 obtidos por HetIC - caso homocedástico com 20% de contaminação.

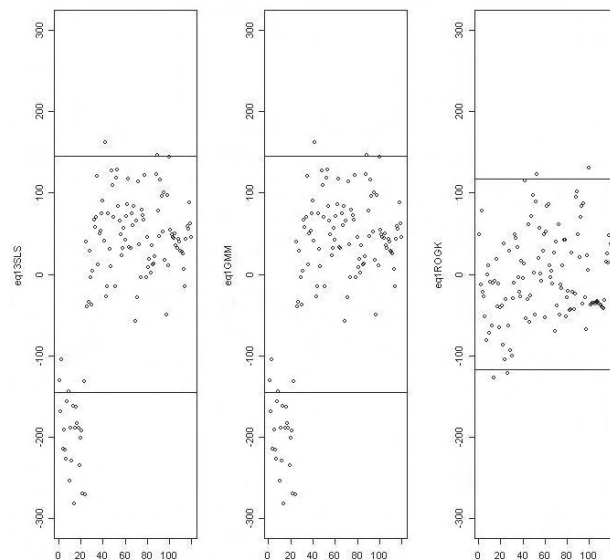


Figura 6.13: Resíduos da equação 1 obtidos por HomIC - caso homocedástico com 20% de contaminação.



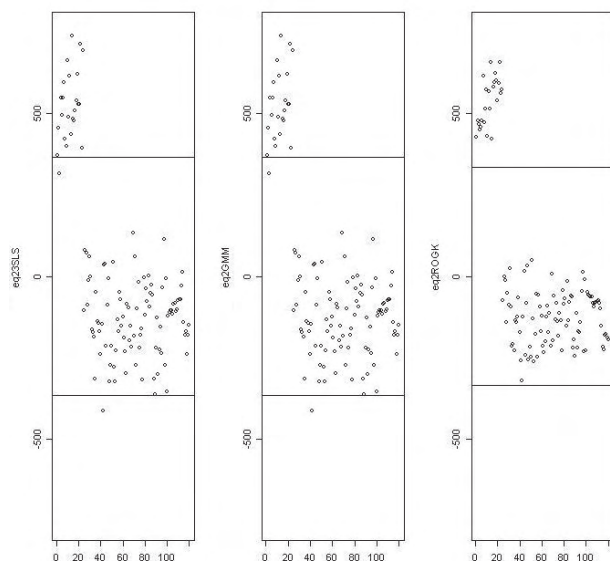


Figura 6.14: Resíduos da equação 2 obtidos por HomIC - caso homocedástico com 20% de contaminação.

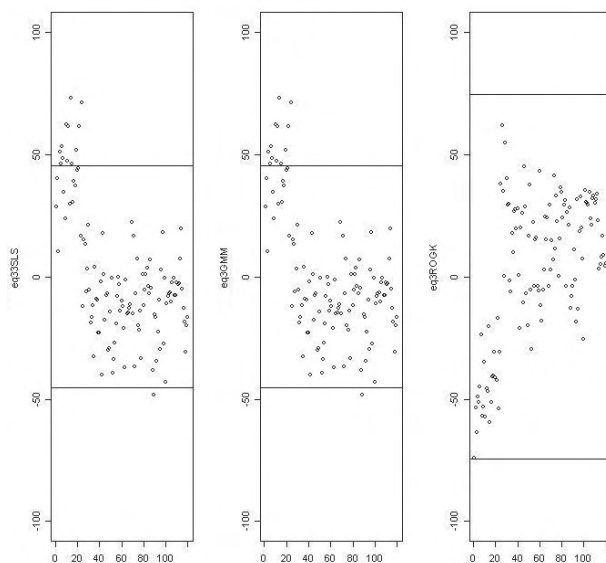


Figura 6.15: Resíduos da equação 3 obtidos por HomIC - caso homocedástico com 20% de contaminação.

caso homocedástico				
	sem contaminação		com 20% de contaminação	
	nvr	$\log(k)$	nvr	$\log(k)$
<i>3SLS</i>	1.5927	0.1779	1.5965	6.1583
<i>GMMIC</i>	1.5927	0.1779	1.5965	6.1583
<i>GMMOGK (HomIC)</i>	1.5932	0.2982	1.6657	4.5370

Tabela 6.5: Médias dos indicadores - caso homocedástico - valores do GMMOGK obtidos com o algoritmo HomIC.

modelo heterocedástico, os resultados obtidos com o estimador *GMMOGK* para os cenários distribucionais NHOC0 e NHOC20 motivaram a que se investigasse melhor o desempenho do estimador nesses casos. Efectivamente, ao construir o estimador *GMMOGK*, prevendo o modelo heterocedástico, existe a necessidade de recorrer à reamostragem para estimar as diversas variâncias dos termos de erro. Assim, decidiu-se estudar o desempenho do estimador quando se admite que o modelo é homocedástico, ou seja, modificando o processo de estimação no sentido de dispensar a reamostragem. Resumindo a ideia, procurou-se estudar o estimador numa versão simplificada e adequada ao caso particular dos modelos homocedásticos. Refizeram-se os cálculos de acordo com a geração das amostras nos cenários NHOC0 e NHOC20, com os indicadores anteriores e com o mesmo número de réplicas.

Os resultados encontrados apresentam-se na Tabela 6.5. A análise desta tabela mostra que o estimador *GMMOGK* apresenta melhores estimativas do que a versão tradicional *GMMIC*, em caso de contaminação. No caso em que não há contaminação os resultados são muito semelhantes. Importa salientar que a adaptação do processo de cálculo das estimativas obtidas por *GMMOGK* ao caso do modelo homocedástico conduz a uma grande simplificação do estimador e a uma redução significativa no tempo de cálculo, uma vez que não envolve qualquer processo de reamostragem.

A representação gráfica dos resíduos por equação, obtidos com a versão simplificada do estimador *GMMOGK* para o caso homocedástico (algoritmo HomIC), para a distribuição homocedástica com 20% de contaminação, correspondem às Figuras 6.13 a 6.15, enquanto que as Figuras 6.16 a 6.18 representam os resultados obtidos pelo mesmo algoritmo para a distribuição homocedástica sem contaminação.

As amplitudes das bandas obtidas com cada método, para cada equação, para o

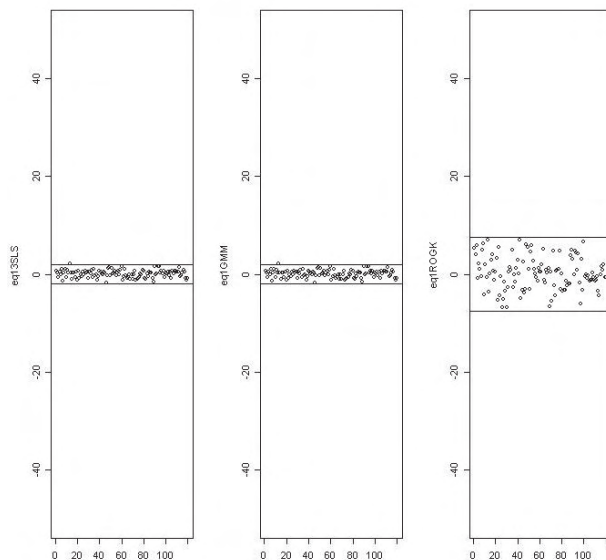


Figura 6.16: Resíduos da equação 1 obtidos com HomIC - caso homocedástico sem contaminação.

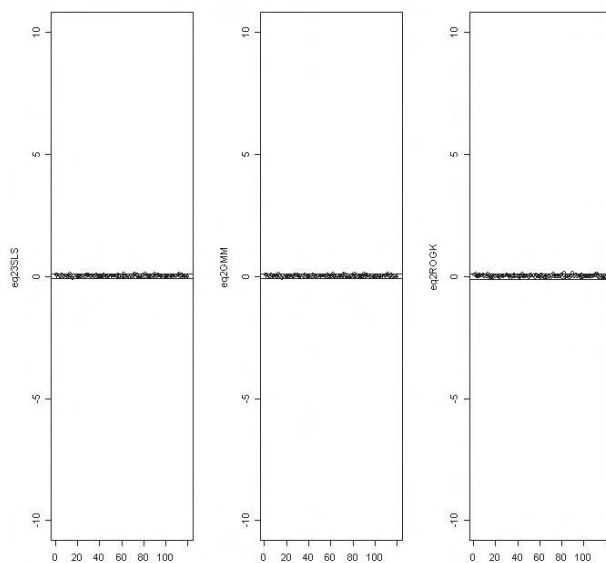


Figura 6.17: Resíduos da equação 2 obtidos com HomIC - caso homocedástico sem contaminação.

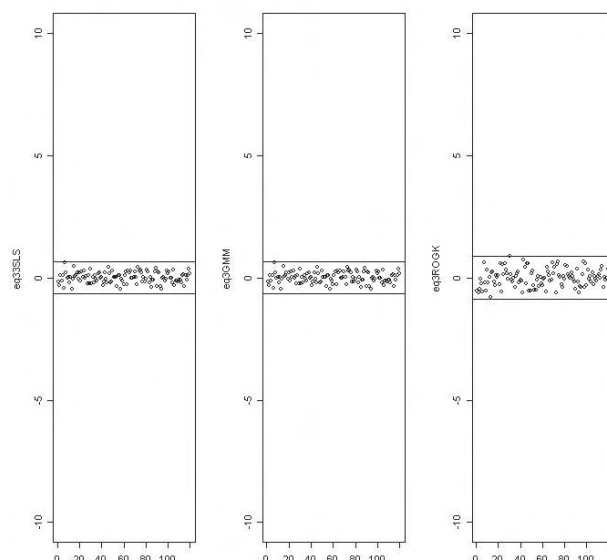


Figura 6.18: Resíduos da equação 3 obtidos com HomIC - caso homocedástico sem contaminação.

caso homocedástico						
	sem contaminação			com 20% de contaminação		
	equação 1	equação 2	equação 3	equação 1	equação 2	equação 3
<i>3SLS</i>	3.8901	0.2088	1.3121	290.2792	730.3481	90.5339
<i>GMMIC</i>	3.8901	0.2088	1.3121	290.2792	730.3481	90.5339
<i>GMMOGK (HomIC)</i>	15.177	0.2207	1.7509	233.7840	670.4168	149.2062

Tabela 6.6: Largura das bandas - caso homocedástico - valores do GMMOGK obtidos com o algoritmo HomIC.

caso homocedástico, sem contaminação e com 20% de contaminação, encontram-se na Tabela 6.6.

Relativamente à última situação amostral considerada,  $T_5 \Sigma$ , ou seja, com distribuições marginais t de Student com 5 graus de liberdade e matriz de covariâncias  $\Sigma$ , obtiveram-se os resultados gráficos que se encontram nas Figuras 6.19 a 6.21, as larguras das bandas constam da Tabela 6.7 e as médias dos indicadores podem ser vistas na Tabela 6.8. Estes resultados são semelhantes aos obtidos nas situações anteriores, variando entre equações e não permitindo destacar nenhum estimador.

Após esta abordagem do estudo e conjugando as conclusões obtidas, é de crer que se

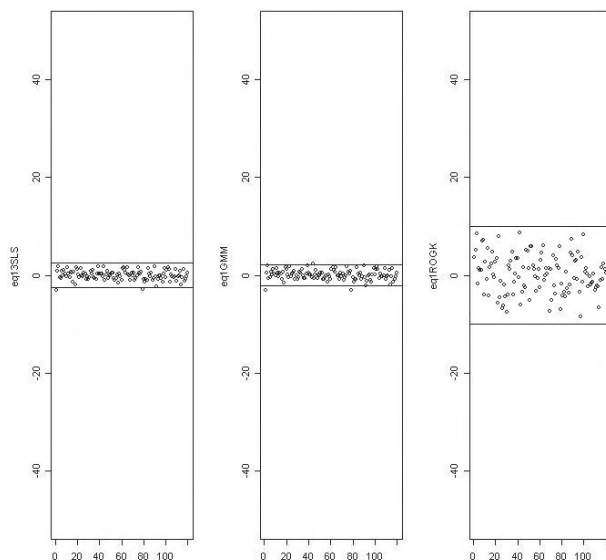


Figura 6.19: Resíduos da equação 1 obtidos com HetIC - caso T5.

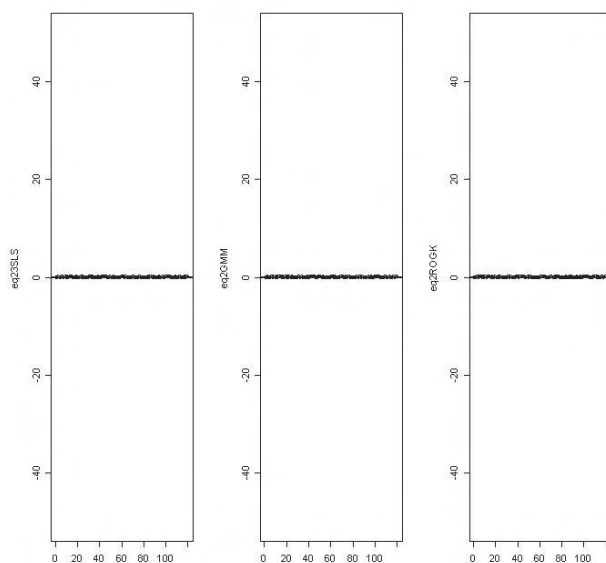


Figura 6.20: Resíduos da equação 2 obtidos com HetIC - caso T5.

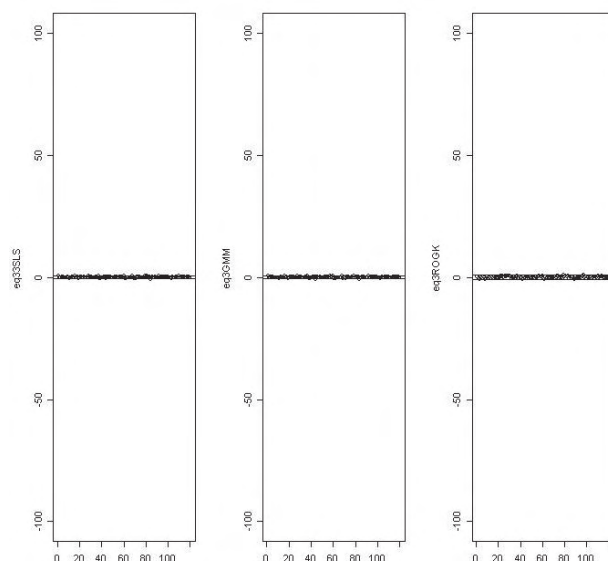


Figura 6.21: Resíduos da equação 3 obtidos com HetIC - caso T5.

	equação 1	equação 2	equação 3
<i>3SLS</i>	4.9184	0.2604	1.0593
<i>GMMIC</i>	4.2926	0.2498	1.1313
<i>GMMOGK (HetIC)</i>	19.9617	0.2712	2.0188

Tabela 6.7: Largura das bandas - caso T5 - valores do GMMOGK obtidos com o algoritmo HetIC.

	nvr	$\log(k)$
<i>3SLS</i>	1.5214	0.2248
<i>GMMIC</i>	1.5255	0.2243
<i>GMMOGK (HetIC)</i>	1.5355	0.3189

Tabela 6.8: Médias dos indicadores - caso T5 - valores do GMMOGK obtidos com o algoritmo HetIC.

justifica usar o estimador *GMMOGK* em situações em que há indícios ou conhecimento concreto de heterocedasticidade. Caso contrário, e apesar de poder haver algum custo na eficiência, poderá valer a pena ignorar a heterocedasticidade do ponto de vista da estimação, usando o *GMMOGK* no seu processo de cálculo simplificado.  $\diamond$

### 6.3 Exemplo proposto por Maronna e Yohai (1997) - dados reais

Este exemplo considera um modelo e um conjunto de dados reais da economia da Argentina, o qual contém observações atípicas. Os dados, referentes ao período de 1956 a 1984, foram apresentados e analisados em Maronna e Yohai (1997) e foram usados no presente trabalho para validar os programas desenvolvidos no decurso deste trabalho. Assim, obtiveram-se estimativas dos parâmetros pelo método dos mínimos quadrados em três fases, as quais foram comparadas com as estimativas publicadas no artigo, verificando-se serem muito semelhantes. Posteriormente, calcularam-se as estimativas dos coeficientes do sistema recorrendo aos estimadores *GMMIC* e *GMMOGK*: apresentam-se e comentam-se os resultados alcançados com a proposta robusta, separadamente para cada uma das equações. Para analisar graficamente os resultados, procedeu-se à representação gráfica dos resíduos por equação, seguindo o procedimento usado por Maronna e Yohai (1997), para facilitar a comparação de resultados. Calcularam-se ainda a norma do vector de resíduos e o logaritmo do número de condição - indicadores sugeridos em Maronna *et al.* (2006). O estudo comparativo evidenciou que a estimação robusta era vantajosa, sobretudo numa das equações.

*Exemplo 6.3.1.* Neste exemplo estudou-se um modelo e um conjunto de dados reais relativos à economia argentina, publicados por Maronna e Yohai (1997). Trata-se de um *SEM* formado por três equações que pode ser apresentado na forma

$$\begin{cases} \mathbf{Y}_1 = & \mathbf{Y}_3\gamma_{31} & + \mathbf{X}_5\beta_{51} + \mathbf{e}_1 \\ \mathbf{Y}_2 = \mathbf{Y}_1\gamma_{12} & + \mathbf{X}_1\beta_{12} & + \mathbf{X}_4\beta_{42} + \mathbf{X}_5\beta_{52} + \mathbf{e}_2 \\ \mathbf{Y}_3 = \mathbf{Y}_1 & - \mathbf{Y}_2 & + \mathbf{X}_1 & + \mathbf{X}_2 - \mathbf{X}_3 + \mathbf{X}_4 \end{cases},$$

ou na forma estrutural,

$$\left\{ \begin{array}{l} -\mathbf{Y}_1 \quad \quad \quad +\mathbf{Y}_3\gamma_{31} \quad \quad \quad \quad \quad \quad \quad \quad +\mathbf{X}_5\beta_{51}+\mathbf{e}_1 \quad = \mathbf{0} \\ \mathbf{Y}_1\gamma_{12}-\mathbf{Y}_2 \quad \quad \quad +\mathbf{X}_1\beta_{12} \quad \quad \quad \quad \quad \quad +\mathbf{X}_4\beta_{42}+\mathbf{X}_5\beta_{52}+\mathbf{e}_2 \quad = \mathbf{0} \\ \mathbf{Y}_1 \quad -\mathbf{Y}_2-\mathbf{Y}_3 \quad \quad \quad +\mathbf{X}_1 \quad \quad \quad +\mathbf{X}_2 - \mathbf{X}_3+\mathbf{X}_4 \quad \quad \quad = \mathbf{0} \end{array} \right. .$$

As variáveis exógenas são:  $x_1$ , o investimento bruto,  $x_2$ , o volume de exportações,  $x_3$ , os custos de impostos e  $x_4$ , o consumo público. Considerou-se uma variável exógena adicional unitária,  $x_5$ , de modo a incluir um termo independente nas equações. Quanto às variáveis endógenas,  $y_1$  representa o consumo privado,  $y_2$  o volume de importações e  $y_3$  o rendimento. O conjunto de dados corresponde a uma matriz com 29 linhas e 7 colunas, contendo valores reais correspondentes aos registos anuais de cada uma das sete variáveis referidas, para o período de 1956 a 1984. Acrescentou-se uma oitava coluna de uns, correspondentes aos valores da variável  $x_5$ . Os termos  $e_1$  e  $e_2$  representam os erros aleatórios do modelo.

Começou-se por estudar a identificabilidade das equações e verificou-se que eram todas identificadas. Para isso, considerou-se a matriz dos parâmetros estruturais do modelo,

$$\Delta_{(8 \times 2)} = \begin{bmatrix} -1 & \gamma_{12} \\ 0 & -1 \\ \gamma_{31} & 0 \\ 0 & \beta_{12} \\ 0 & 0 \\ 0 & 0 \\ 0 & \beta_{42} \\ \beta_{51} & \beta_{52} \end{bmatrix},$$

construiu-se a matriz dos coeficientes das restrições de exclusão, para cada equação  $i$  na forma estrutural, a qual se designou por  $\mathbf{R}_i$ , e discutiu-se a identificabilidade, de acordo com a **Proposição 2.3.4**.



Assim, para a primeira equação, tem-se que

$$\mathbf{R}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\Rightarrow \mathbf{R}_1 \mathbf{\Delta} = \begin{bmatrix} 0 & -1 \\ 0 & \beta_{12} \\ 0 & 0 \\ 0 & 0 \\ 0 & \beta_{42} \end{bmatrix},$$

logo  $rank(\mathbf{R}_1 \mathbf{\Delta}) = 1$  e  $rank(\mathbf{R}_1) = 5 > 1$ , concluindo-se que esta equação é sobreidentificada.

Quanto à segunda equação, tem-se que

$$\mathbf{R}_2 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$\Rightarrow \mathbf{R}_2 \mathbf{\Delta} = \begin{bmatrix} \gamma_{31} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

pelo que  $rank(\mathbf{R}_2 \mathbf{\Delta}) = 1$  e  $rank(\mathbf{R}_2) = 3 > 1$ , donde se conclui que esta equação também é sobreidentificada.

Na terceira equação, tem-se

$$\mathbf{R}_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\Rightarrow \mathbf{R}_3 \mathbf{\Delta} = \begin{bmatrix} \beta_{51} & \beta_{52} \end{bmatrix},$$

donde se conclui que  $rank(\mathbf{R}_3 \mathbf{\Delta}) = 1$  e  $rank(\mathbf{R}_3) = 1 = 1$ , logo esta equação é exactamente identificada.

Considerando o modelo na forma  $\mathbf{y} = \mathbf{Z}\boldsymbol{\delta} + \mathbf{e}$ , o vector de parâmetros estruturais desconhecidos que interessa estimar é

$$\boldsymbol{\delta} = \begin{bmatrix} \gamma_{31} & \beta_{51} & \gamma_{12} & \beta_{12} & \beta_{42} & \beta_{52} \end{bmatrix}'.$$

Note-se que a última equação é uma identidade contabilística, não contém parâmetros desconhecidos, logo não necessita de ser estimada.

No trabalho de Maronna e Yohai (1997), os autores apresentam os resultados das estimativas do vector de parâmetros obtidas pelo método dos mínimos quadrados iterados em três fases e por um estimador robusto proposto, o estimador  $\tau$  multivariado, os quais podem ver-se a seguir.

$$\hat{\delta}_{I3SLS} = \begin{bmatrix} 0.676 & 134.89 & 0.449 & -0.370 & -0.691 & -125.121 \end{bmatrix}' \text{ e} \\ \hat{\delta}_{\tau} = \begin{bmatrix} 0.728 & 80.438 & 0.124 & 0.001 & 0.088 & 0.765 \end{bmatrix}'.$$

Esses autores efectuaram a análise dos resultados individualmente, para cada uma das equações do modelo. Para comparar os resultados obtidos pelos diferentes métodos, basearam-se nos valores dos resíduos e estudaram a distribuição empírica dos resíduos por equação. A avaliação das estimativas foi feita tendo em conta a largura das bandas definidas por  $2 \times 2.5 \times MAD$ .

No presente trabalho, estimaram-se os coeficientes estruturais do modelo pelos diferentes métodos que foram referidos anteriormente, isto é, por mínimos quadrados em três fases, por momentos generalizado e pela versão robusta dos momentos generalizado. Na análise dos resultados, utilizaram-se os critérios usados por Maronna e Yohai (1997), de modo a facilitar a comparação. Foram obtidas estimativas do vector  $\delta$  e os resultados foram os seguintes:

$$\hat{\delta}_{3SLS} = \begin{bmatrix} 0.676 & 135.32 & 0.446 & -0.367 & -0.680 & -124.471 \end{bmatrix}', \\ \hat{\delta}_{GMMIC} = \begin{bmatrix} 0.673 & 136.27 & 0.599 & -0.651 & -1.257 & -136.884 \end{bmatrix}' \text{ e} \\ \hat{\delta}_{GMMOGK} = \begin{bmatrix} 0.686 & 120.05 & 0.456 & -0.385 & -0.657 & -131.871 \end{bmatrix}'.$$

Relativamente à estimativa obtida por  $3SLS$ , que foi calculada com o objectivo de controlo, observou-se ser muito próxima da obtida por Maronna e Yohai (1997) por  $I3SLS$ . Para analisar os resultados por equação, representaram-se separadamente os resíduos, obtidos por  $GMMIC$  e por  $GMMOGK$ , correspondentes a cada uma das equações e duas bandas definidas pelos limites por  $l_1 = -2.5 \times MAD$  e  $l_2 = +2.5 \times MAD$ . Os resultados gráficos encontram-se nas Figuras 6.16 e 6.17.

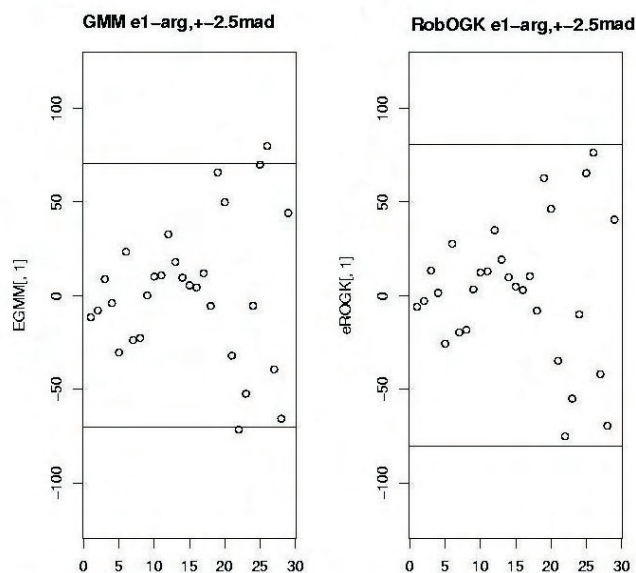


Figura 6.22: Resíduos obtidos por *GMMIC* e por *GMMOGK* para a 1ª equação - dados da Argentina.

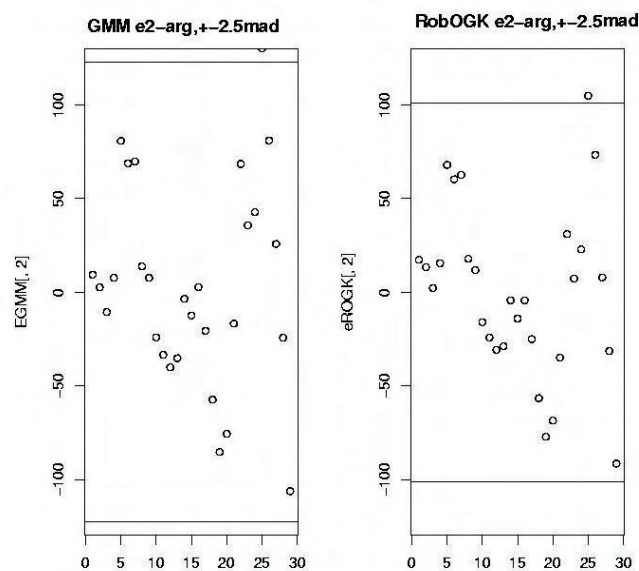


Figura 6.23: Resíduos obtidos por *GMMIC* e por *GMMOGK* para a 2ª equação - dados da Argentina.

	equação 1	equação 2
<i>GMMIC</i>	140.9019	245.2970
<i>GMMOGK (HetIC)</i>	159.8163	193.4912

Tabela 6.9: Largura das bandas obtidas - dados da Argentina.

	nvr	$\log(k)$
<i>GMMIC</i>	1.11	0.41
<i>GMMOGK (HetIC)</i>	1.13	0.31

Tabela 6.10: Valores dos indicadores - dados da Argentina.

Comparando os resultados obtidos pela versão tradicional do estimador *GMM*, com os da versão robusta, confirmou-se que as conclusões diferem de equação para equação, tal como havia sido verificado no artigo anteriormente referido. A Tabela 6.9 contém as amplitudes das bandas obtidas com cada método, para cada equação.

Os valores encontrados indicam que, no caso da segunda equação, a versão robusta do estimador *GMM* permite obter melhores resultados em relação à versão clássica do estimador *GMM*, uma vez que os resíduos correspondentes à versão robusta apresentam menor dispersão, quando esta foi avaliada pela da largura das bandas. Por outro lado, a Tabela 6.10 contém os valores dos indicadores sugeridos por Maronna *et al.* (2006) (norma do vector de resíduos e logaritmo o número de condição), obtidos com cada uma das versões do estimador *GMM*.

Avaliando a qualidade do *GMMOGK* com base na distribuição dos resíduos, verifica-se que as estimativas dos parâmetros são próximas das que se se obtêm por outros métodos e há um ganho significativo em termos de dispersão dos resíduos.

Esta ilustração mostrou que os resultados foram claramente satisfatórios. Apesar da metodologia robusta, os resultados encontrados parecem indicar que as conclusões diferem de equação para equação.  $\diamond$

# Capítulo 7

## Conclusões

### 7.1 Comentários gerais

O presente trabalho começa por destacar os papéis que tanto a Economia como a Econometria e a Estatística desempenham na concepção e na análise do Modelo de Equações Simultâneas. Embora o ponto de vista/interesse deste estudo seja essencialmente estatístico, achou-se interessante situar as origens do modelo e as suas características econométricas, aspectos bem divulgados em obras de econometria, mas não nas de estatística, onde o próprio modelo é geralmente ignorado.

De seguida, apresentou-se a caracterização formal dos modelos e das condições distribucionais que, em geral, são assumidas. Para além disso, fez-se uma breve revisão do estudo dos modelos, os quais possuem características específicas, decorrentes da motivação económica subjacente, mas que exigem um tratamento matemático considerável.

Admitindo que o sistema de equações verifica as condições algébricas necessárias à identificação dos parâmetros estatísticos, que incluem os coeficientes das variáveis económicas, passou-se a descrever os processos de estimação mais divulgados, nomeadamente, a estimação dos parâmetros pelo método dos mínimos quadrados e pelo método dos momentos generalizado. O estudo focalizou-se no método dos momentos generalizado, tendo em conta que, na literatura mais recente, este último método tem sido preferido, por permitir integrar no processo de construção dos estimadores condições de optimalidade decorrentes da situação económica em análise. Por outro lado, a utilização do método dos momentos generalizado não exige que o modelo seja homocedástico,

o que permitiu estudar o processo de estimação na forma mais geral do modelo heterodástico.

Uma das questões específicas do modelo de equações simultâneas é a polémica que existe entre a opção pela estimação dos parâmetros através da estimação sucessiva de cada uma das equações do sistema (cujos métodos são designados globalmente por métodos de informação limitada) ou, pelo contrário, pela estimação conjunta de todas as equações (sendo estes métodos conhecidos por métodos de informação completa). Os estudos publicados sobre este assunto comparam as duas vias em termos de eficiência dos estimadores (supondo que os erros do modelo seguem distribuição normal), mas apenas estudam os métodos baseados em estimadores de mínimos quadrados. Entretanto, alguns autores continuam a apontar as vantagens empíricas dos métodos de informação limitada. Uma vez que a questão não tem sido investigada para os estimadores obtidos pelo método dos momentos generalizado, este trabalho incluiu um estudo comparativo original, desenvolvido nesta classe de estimadores. Após ter sido encontrada uma representação conveniente para a matriz de covariâncias dos erros do modelo, foi possível demonstrar que os estimadores de momentos generalizados são assintoticamente mais eficientes usando informação completa do que quando se usam métodos de informação limitada. O resultado foi estabelecido com base no critério da variância generalizada. Para além disso, a referida representação da matriz de covariâncias dos erros permitiu obter uma forma explícita para o estimador de momentos generalizados, em função de uma estimativa dessa matriz. Posteriormente, tomou-se conhecimento de que esta forma do estimador tinha já sido desenvolvida e publicada por outros autores, embora fosse apresentada através de uma expressão bem diferente. Este facto, temporariamente desagradável, tornou-se útil para confirmar a validade do resultado encontrado por uma via original.

Como a demonstração efectuada sobre a eficiência relativa dos estimadores é válida apenas em termos assintóticos, ampliou-se a investigação do problema, considerando um estudo paralelo, empírico, com amostras de dimensão finita. Esse estudo incidiu sobre dois modelos particulares e verificou-se que, em ambos os casos, os resultados obtidos eram idênticos pelos dois processos (informação completa ou informação limitada). Assim sendo, e tendo em conta as vantagens da estimação limitada, quer pela

maior simplicidade no tratamento matemático, quer pela menor propagação de erros de modelação, concluiu-se que, pelo menos para os modelos particulares considerados, em amostras de dimensão finita não há vantagens práticas em optar pela estimação simultânea de todas as equações do sistema. Esta conclusão confirma que há motivos justificados para que alguns autores (por exemplo, Greene, 2003 e Baltagi, 2008) recomendem que sejam usados métodos de informação limitada em situações reais.

De acordo com as conclusões teóricas atrás referidas, tendo em conta o objetivo principal do trabalho, que consistia na pesquisa de estimadores robustos para os parâmetros do modelo de equações simultâneas, e apesar das reflexões sobre a estimação em amostras de dimensão finita, foi decidido prosseguir a investigação considerando os cenários mais gerais do modelo, ou seja, supondo o modelo heterocedástico e usando a metodologia da informação completa. Esta decisão teve em conta as seguintes conveniências: i) do geral (modelo heterocedástico) decorre o particular (modelo homocedástico), ii) a elegância matemática, probabilística e estatística é maior no tratamento da estimação simultânea, iii) o pesado trabalho computacional, que no passado era uma desvantagem significativa, é agora executado rapidamente pela maior parte dos computadores portáteis.

O trabalho prosseguiu então com a procura de um processo de estimação constituído por três etapas (tal como acontece na estimação tradicional pelo método dos momentos generalizado) que conduzisse a um estimador global robusto. A linha orientadora da investigação consistiu em modificar cada uma das três fases, substituindo ou adaptando os estimadores tradicionais intervenientes em cada fase, por estimadores robustos. Tendo em atenção as características comuns entre os modelos de equações simultâneas e os modelos de regressão multivariada, estudou-se com maior detalhe a eventual adequação de métodos da regressão robusta. Por outro lado, a representação explícita do estimador como função da estimativa da matriz das covariâncias dos erros, obrigou à selecção do melhor método, de entre as diversas propostas existentes para a estimação robusta de matrizes de covariâncias.

Deste modo, foi possível desenvolver a proposta de uma versão robusta do estimador *GMM*, que se designou por *GMMOGK*, e que, resumidamente, consiste nas seguintes fases sucessivas:

1. *Cálculo de estimativas iniciais dos parâmetros.* Tal como no processo tradicional, as estimativas iniciais dos resíduos foram obtidas separadamente para cada equação do modelo, mas agora usando o estimador robusto *LTS*. É de referir que o estimador *LTS* conduziu sempre a excelentes soluções.
2. *Adaptação aos SEM do método ortogonalizado de Gnanadesikan - Kettenring.* Para estimar a matriz de covariâncias dos erros do modelo foi necessário efectuar previamente a reamostragem dos resíduos obtidos na fase 1. Ao recorrer à reamostragem é inevitável um aumento significativo do tempo computacional no cálculo das estimativas. Mesmo assim, o tempo de cálculo é curto, ficando esta opção perfeitamente justificada.
3. *Estimação dos coeficientes estruturais.* Foi efectuada através da minimização da função de Huber, depois de ponderada com as estimativas "robustas" encontradas na fase anterior. A opção feita foi condicionada pela heterocedasticidade e pela dependência entre equações.

Para avaliar o desempenho do estimador robusto proposto, fizeram-se estudos de aplicação de natureza distinta, sempre recorrendo ao programa *R* (e às suas *packages*) para efectuar todos os cálculos:

- i) Estudos de simulação - neste caso, as análises contemplaram, entre outras, as seguintes distribuições: normal heterocedástica sem contaminação, normal heterocedástica com 20% de contaminação, normal homocedástica sem contaminação, normal homocedástica com 20% de contaminação e distribuições marginais *t* de Student com correlacionamento entre componentes. Para confirmar os resultados procedeu-se à replicação das simulações efectuadas, o que obrigou a algumas adaptações do método, em situações de heterocedasticidade. Para além das situações descritas na tese, foi efectuada uma grande diversidade de simulações que não foram incluídas, por se entender que não conduziam a material relevante em relação ao exposto.
- ii) Análise de dados reais - analisou-se um conjunto de dados relativos à economia argentina; tratou-se de um estudo essencialmente comparativo, onde foi possível



usar as observações reais para comparar as estimativas agora calculadas com as estimativas anteriormente obtidas com outros métodos (publicadas em Maronna e Yohai, 1997).

Os critérios de avaliação basearam-se no estudo da distribuição empírica dos resíduos multivariados, tendo em conta dois indicadores de natureza numérica e um de natureza gráfica: i) usaram-se dois indicadores numéricos, propostos em Maronna *et al.* (2006), um relativo à localização (norma dos vectores de resíduos) e outro relativo à escala (logaritmo do número de condição da matriz de covariâncias dos resíduos), ii) o critério gráfico consistiu na representação, em cada equação, das medianas dos resíduos e de duas bandas que facilitaram a avaliação da dispersão dos resíduos.

A análise efectuada sobre os resultados destes indicadores permite concluir que a versão robusta *GMMOGK*, desenvolvida para o modelo heterocedástico, apresenta um melhor desempenho precisamente quando os dados são heterocedásticos. Com efeito, nos estudos de simulação efectuados, observou-se que se esse estimador fosse aplicado a dados provenientes de um modelo homocedástico, os resultados não eram favoráveis. No entanto, ao aplicar a versão simplificada do *GMMOGK* para o caso particular homocedástico, a dados gerados por um modelo homocedástico, já foram obtidos bons resultados. Neste último caso, na presença de contaminação, as estimativas foram claramente melhores do que as obtidas pelo *GMM* tradicional; sem contaminação os resultados foram próximos dos obtidos com os métodos tradicionais. Com dados reais, desconhece-se se a amostra corresponde a um modelo heterocedástico, ou não. Assim, se não existirem indícios significativos de heterocedasticidade, parece ser preferível usar a versão simplificada do *GMMOGK*, pois tem vantagens do ponto de vista da robustez e de tempo de cálculo, uma vez que evita a fase de reamostragem dos resíduos.

Finalmente, importa referir que, independentemente da metodologia usada ser robusta ou tradicional, se os resultados forem analisados separadamente para cada equação, o estudo efectuado parece indicar que a qualidade das estimativas depende sempre da estrutura de cada equação. Efectivamente, o *GMMOGK* foi concebido para a estimação simultânea de todas as equações, mas os resultados podem também ser analisados equação a equação. Caso seja feita essa análise, verifica-se que as conclusões não são idênticas para todas as equações do sistema. De qualquer modo, é de realçar

que este facto também ocorre com os estimadores tradicionais, pelo que não se pode considerar que seja uma desvantagem do método proposto nesta tese.

## 7.2 Perspectivas de trabalho futuro

O trabalho efectuado permitiu confirmar que em face do interesse que existe na aplicação do *SEM*, se justifica que sejam investigados muitos aspectos ainda em aberto, tanto sobre o próprio modelo, como do ponto de vista particular da robustez dos estimadores.

A título de exemplo, apontam-se algumas questões que se identificaram como tendo interesse estudar, mas que ficaram por explorar:

- A comparação das matrizes de covariâncias assintóticas dos estimadores *GMM* com informação completa e com informação limitada foi efectuada com base no critério da variância generalizada. Para além do estudo deste assunto que se apresenta na tese, tentou-se estabelecer a comparação das matrizes com base num critério mais forte. Lamentavelmente, para este último critério, não foi possível alcançar resultados análogos aos que se apresentam.
- O estudo da aplicação do *GMMOGK* de informação limitada. Com efeito, neste trabalho optou-se por estudar o caso mais geral (e trabalhoso) da aplicação do método de informação completa. Atendendo à controvérsia existente entre a opção por estimadores de informação limitada ou de informação completa, à superioridade pouco evidente da informação completa em dimensão finita, considera-se que este tópico merece um investigação aprofundada.
- O grande número de parâmetros envolvidos no sistema de equações de um *SEM* e a natureza das distribuições, com variáveis multidimensionais e dependentes, são causas de grande complexidade na simulação de cenários de contaminação. A grande quantidade de hipóteses de contaminação que foram simuladas na preparação deste trabalho, mas não descritas no texto, fez sentir a necessidade de averiguar quais os casos de contaminação que mais afectam os estimadores tradicionais.

- Verificou-se que, em todos os exemplos estudados (quer publicados noutros trabalhos, quer analisados na presente tese), não foi possível identificar, em cada equação, quais são os elementos decisivos na qualidade das estimativas.
- Foram infrutíferas as tentativas para encontrar um *SEM* que tenha sido proposto para a economia portuguesa. Entende-se que seria interessante conhecer tais modelos e correspondentes dados reais, de modo a estudar o desempenho do estimador *GMMOGK* na estimação dos parâmetros. Eventualmente, esta tarefa será possível futuramente, num projecto conjunto com colegas de Econometria.
- Do ponto de vista computacional, foi excelente a facilidade e a possibilidade de ter acesso ao programa *R*, de forma a poder trabalhar com o programa e com alguma das suas *packages*. Infelizmente, para calcular as estimativas *GMMOGK* não foi possível usar directamente nenhuma das *packages* disponíveis (em particular devido à heterocedasticidade e à reamostragem necessária para a aplicação do estimador *OGK* aos *SEM*). Seria muito útil desenvolver um projecto com colegas da estatística computacional, de modo a otimizar as macros escritas pela autora e a permitir disponibilizar numa *package* o processo de cálculo das estimativas pelo estimador robusto proposto.
- Finalmente, espera-se que o presente trabalho possa ter dado alguma contribuição para o estudo estatístico dos *SEM*, e que possa sugerir novas investigações a outros autores.

# Apêndices

A.1.1 Reescrevendo (2.2.1) na forma matricial, em relação à observação  $t, t = 1, \dots, T$ ,

vem

$$\begin{bmatrix} \mathbf{y}_1 & \cdots & \mathbf{y}_M \end{bmatrix}_t \begin{bmatrix} \nu_{11} & \cdots & \nu_{1M} \\ \vdots & & \vdots \\ \nu_{M1} & \cdots & \nu_{MM} \end{bmatrix} + \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_k \end{bmatrix}_t \begin{bmatrix} \beta_{11} & \cdots & \beta_{1M} \\ \vdots & & \vdots \\ \beta_{k1} & \cdots & \beta_{kM} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_M \end{bmatrix}_t = \mathbf{0}_{(1 \times M)}.$$

Relativamente à forma matricial estrutural (2.2.2), tem-se

$$\begin{aligned} \mathbf{Y}_{(T \times M)} &= [y_{ti}]_{t=1, \dots, T, i=1, \dots, M} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_1 & \cdots & \mathbf{Y}_T \end{bmatrix}', \\ \mathbf{X}_{(T \times k)} &= [x_{tj}]_{t=1, \dots, T, j=1, \dots, k} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_k \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_T \end{bmatrix}', \\ \mathbf{E}_{(T \times M)} &= [e_{ti}]_{t=1, \dots, T, i=1, \dots, M} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_M \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_T \end{bmatrix}', \end{aligned}$$

onde  $\mathbf{Y}'_t$ ,  $\mathbf{x}'_t$  e  $\mathbf{e}'_t$  representam a linha  $t$  das matrizes  $\mathbf{Y}$ ,  $\mathbf{X}$  e  $\mathbf{E}$ , respectivamente, e representando por  $\mathbf{A}'$  a transposta da matriz  $\mathbf{A}$ .

A.1.2 A partir de (2.2.2) é possível exprimir o sistema para cada observação  $t$ , em termos das matrizes dos coeficientes.

$$\begin{aligned} & \begin{bmatrix} y_{11} & \cdots & y_{1M} \\ \vdots & & \vdots \\ y_{T1} & \cdots & y_{TM} \end{bmatrix} \begin{bmatrix} \nu_{11} & \cdots & \nu_{1M} \\ \vdots & & \vdots \\ \nu_{M1} & \cdots & \nu_{MM} \end{bmatrix} + \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & & \vdots \\ x_{T1} & \cdots & x_{Tk} \end{bmatrix} \begin{bmatrix} \beta_{11} & \cdots & \beta_{1M} \\ \vdots & & \vdots \\ \beta_{k1} & \cdots & \beta_{kM} \end{bmatrix} + \\ & + \begin{bmatrix} e_{11} & \cdots & e_{1M} \\ \vdots & & \vdots \\ e_{T1} & \cdots & e_{TM} \end{bmatrix} = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix}_{(T \times M)} \end{aligned}$$

$$\begin{aligned}
&\Leftrightarrow \begin{bmatrix} \sum_{i=1}^M y_{1i}\nu_{i1} & \cdots & \sum_{i=1}^M y_{1i}\nu_{iM} \\ \vdots & & \vdots \\ \sum_{i=1}^M y_{Ti}\nu_{i1} & \cdots & \sum_{i=1}^M y_{Ti}\nu_{iM} \end{bmatrix} + \begin{bmatrix} \sum_{j=1}^k x_{1j}\beta_{j1} & \cdots & \sum_{j=1}^k x_{1j}\beta_{jM} \\ \vdots & & \vdots \\ \sum_{j=1}^k x_{Tj}\beta_{j1} & \cdots & \sum_{j=1}^k x_{Tj}\beta_{jM} \end{bmatrix} + \\
&+ \begin{bmatrix} e_{11} & \cdots & e_{1M} \\ \vdots & & \vdots \\ e_{T1} & \cdots & e_{TM} \end{bmatrix} = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix}_{(T \times M)},
\end{aligned}$$

pelo que o modelo (2.2.2) é equivalente a ter

$$\begin{bmatrix} \sum_{i=1}^M y_{1i}\nu_{i1} + \sum_{j=1}^k x_{1j}\beta_{j1} + e_{11} = 0 & \cdots & \sum_{i=1}^M y_{1i}\nu_{iM} + \sum_{j=1}^k x_{1j}\beta_{jM} + e_{1M} = 0 \\ \vdots & & \vdots \\ \sum_{i=1}^M y_{Ti}\nu_{i1} + \sum_{j=1}^k x_{Tj}\beta_{j1} + e_{T1} = 0 & \cdots & \sum_{i=1}^M y_{Ti}\nu_{iM} + \sum_{j=1}^k x_{Tj}\beta_{jM} + e_{TM} = 0 \end{bmatrix}, \quad (7.2.1)$$

podendo esta forma ser interpretada do seguinte modo: as  $M$  componentes da linha  $t$  correspondem à  $t$  éxima observação de cada uma das  $M$  equações, enquanto que as  $T$  componentes da coluna  $i$  correspondem às  $T$  observações da equação  $i$ .

- (a) Observe-se que, fixando a observação  $t$  do sistema de  $M$  equações na forma (7.2.1), pode escrever-se

$$\mathbf{Y}'_t \boldsymbol{\Gamma} + \mathbf{x}'_t \mathbf{B} + \mathbf{e}'_t = \mathbf{0}_{(1 \times M)}, t = 1, \dots, T,$$

onde  $\mathbf{Y}'_t$ ,  $\mathbf{x}'_t$  e  $\mathbf{e}'_t$  representam a linha  $t$  das matrizes  $\mathbf{Y}$ ,  $\mathbf{X}$  e  $\mathbf{E}$ , respectivamente.

- (b) Por outro lado, destacando a equação  $i$  em (7.2.1) e considerando o conjunto das  $T$  observações, vem

$$\begin{cases} \sum_{g=1}^M y_{1g}\nu_{gi} + \sum_{j=1}^k x_{1j}\beta_{ji} + e_{1i} = 0 \\ \vdots \\ \sum_{g=1}^M y_{Tg}\nu_{gi} + \sum_{j=1}^k x_{Tj}\beta_{ji} + e_{Ti} = 0 \end{cases}$$

$$\Leftrightarrow \mathbf{Y} \boldsymbol{\Gamma}_i + \mathbf{X} \mathbf{B}_i + \mathbf{e}_i = \mathbf{0}_{(T \times 1)}, i = 1, \dots, M,$$

onde  $\boldsymbol{\Gamma}_i$ ,  $\mathbf{B}_i$  e  $\mathbf{e}_i$  representam a coluna  $i$  das matrizes  $\boldsymbol{\Gamma}$ ,  $\mathbf{B}$  e  $\mathbf{E}$ , respectivamente.

A.1.3 Designou-se por  $\mathring{\Sigma}$  a matriz  $E \left[ \frac{\mathbf{E}'\mathbf{E}}{T} | \mathbf{X} \right]$ . Tendo em conta HFE.1.3-(i), esta matriz pode tomar a forma seguinte:

$$\begin{aligned} \mathring{\Sigma} &= E \left[ \frac{\mathbf{E}'\mathbf{E}}{T} | \mathbf{X} \right] \\ &= E \left[ \frac{1}{T} \begin{pmatrix} e_{11} & \cdots & e_{T1} \\ \vdots & & \vdots \\ e_{1M} & \cdots & e_{TM} \end{pmatrix} \begin{pmatrix} e_{11} & \cdots & e_{1M} \\ \vdots & & \vdots \\ e_{T1} & \cdots & e_{TM} \end{pmatrix} \right] \\ &= E \left[ \frac{1}{T} \begin{pmatrix} \sum_{t=1}^T e_{t1}e_{t1} & \cdots & \sum_{t=1}^T e_{t1}e_{tM} \\ \vdots & & \vdots \\ \sum_{t=1}^T e_{tM}e_{t1} & \cdots & \sum_{t=1}^T e_{tM}e_{tM} \end{pmatrix} \right] \\ &= \begin{bmatrix} \frac{1}{T} \sum_{t=1}^T \sigma_{1t}^2 & \cdots & \frac{1}{T} \sum_{t=1}^T \sigma_{1Mt} \\ \vdots & \ddots & \vdots \\ \frac{1}{T} \sum_{t=1}^T \sigma_{Mt} & \cdots & \frac{1}{T} \sum_{t=1}^T \sigma_{Mt}^2 \end{bmatrix}. \end{aligned}$$

Repare-se que, no caso homocedástico, quando se verifica  $[\sigma_{it}^2 = \sigma_i^2 \wedge \sigma_{ijt} = \sigma_{ij}], \forall t = 1, \dots, T$ , esta matriz se simplifica, permitindo obter a forma

$$\begin{bmatrix} \frac{1}{T} (T\sigma_1^2) & \cdots & \frac{1}{T} (T\sigma_{1M}) \\ \vdots & \ddots & \vdots \\ \frac{1}{T} (T\sigma_{M1}) & \cdots & \frac{1}{T} (T\sigma_M^2) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1M} \\ \vdots & \ddots & \vdots \\ \sigma_{M1} & \cdots & \sigma_M^2 \end{bmatrix} = \Sigma.$$

A.2.1 Propriedades do estimador  $\hat{\delta}_{i,OLS}$ .

POLS.1 O estimador  $\hat{\delta}_{i,OLS}$  não é consistente para  $\delta_i$ .

*Demonstração.* Para verificar a consistência do estimador, note-se que

$$p \lim_{T \rightarrow \infty} [\hat{\delta}_{i,OLS}] = \delta_i + p \lim_{T \rightarrow \infty} \left[ \frac{\frac{\mathbf{Y}_i' \mathbf{Y}_i}{T}}{\frac{\mathbf{X}_i' \mathbf{Y}_i}{T}} \frac{\frac{\mathbf{Y}_i' \mathbf{X}_i}{T}}{\frac{\mathbf{X}_i' \mathbf{X}_i}{T}} \right]^{-1} \times p \lim_{T \rightarrow \infty} \left[ \frac{\frac{\mathbf{Y}_i' \mathbf{e}_i}{T}}{\frac{\mathbf{X}_i' \mathbf{e}_i}{T}} \right]. \quad (7.2.2)$$

Para estudar estes limites, importa escrever a forma reduzida de acordo com a partição seguida para a forma estrutural, em (2.3.2), obtendo-se:

$$\begin{bmatrix} \mathbf{y}_i & \mathbf{Y}_i & \mathbf{Y}_i^* \end{bmatrix} = \begin{bmatrix} \mathbf{X}_i & \mathbf{X}_i^* \end{bmatrix} \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \end{bmatrix} + \begin{bmatrix} \mathbf{v}_i & \mathbf{V}_i & \mathbf{V}_i^* \end{bmatrix}. \quad (7.2.3)$$

Por outro lado, tendo em conta a **Proposição 2.2.3**, tem-se

$$p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{Y}'\mathbf{Y}) = \mathbf{\Pi}' \mathbf{\Sigma}_{\mathbf{XX}} \mathbf{\Pi} + \mathbf{\Sigma}^v,$$

onde

$$\begin{aligned} \mathbf{\Sigma}_{\mathbf{XX}} &= p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{X}'\mathbf{X}) \Leftrightarrow \mathbf{\Sigma}_{\mathbf{XX}} = p \lim_{T \rightarrow \infty} \frac{1}{T} \left\{ \begin{bmatrix} \mathbf{X}'_i \\ \mathbf{X}_i^{*'} \end{bmatrix} \begin{bmatrix} \mathbf{X}_i & \mathbf{X}_i^* \end{bmatrix} \right\} \\ &\Leftrightarrow \mathbf{\Sigma}_{\mathbf{XX}} = p \lim_{T \rightarrow \infty} \frac{1}{T} \begin{bmatrix} \mathbf{X}'_i \mathbf{X}_i & \mathbf{X}'_i \mathbf{X}_i^* \\ \mathbf{X}_i^{*'} \mathbf{X}_i & \mathbf{X}_i^{*'} \mathbf{X}_i^* \end{bmatrix} \end{aligned} \quad (7.2.4)$$

$$\Rightarrow \mathbf{\Sigma}_{\mathbf{XX}} = \begin{bmatrix} (\mathbf{\Sigma}_{\mathbf{XX}})_{11} & (\mathbf{\Sigma}_{\mathbf{XX}})_{12} \\ (\mathbf{\Sigma}_{\mathbf{XX}})_{21} & (\mathbf{\Sigma}_{\mathbf{XX}})_{22} \end{bmatrix} = \begin{bmatrix} (\mathbf{\Sigma}_{\mathbf{XX}})_{.1} & (\mathbf{\Sigma}_{\mathbf{XX}})_{.2} \end{bmatrix}. \quad (7.2.5)$$

Assim, pode escrever-se

$$\begin{aligned} p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{Y}'\mathbf{Y}) &= \begin{bmatrix} \pi'_{11} & \pi'_{21} \\ \pi'_{12} & \pi'_{22} \\ \pi'_{13} & \pi'_{23} \end{bmatrix} \begin{bmatrix} (\mathbf{\Sigma}_{\mathbf{XX}})_{11} & (\mathbf{\Sigma}_{\mathbf{XX}})_{12} \\ (\mathbf{\Sigma}_{\mathbf{XX}})_{21} & (\mathbf{\Sigma}_{\mathbf{XX}})_{22} \end{bmatrix} \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \end{bmatrix} + \mathbf{\Sigma}^v, \\ &\Leftrightarrow p \lim_{T \rightarrow \infty} \frac{1}{T} \begin{bmatrix} \mathbf{y}'_i \mathbf{y}_i & \mathbf{y}'_i \mathbf{Y}_i & \mathbf{y}'_i \mathbf{Y}_i^* \\ \mathbf{Y}'_i \mathbf{y}_i & \mathbf{Y}'_i \mathbf{Y}_i & \mathbf{Y}'_i \mathbf{Y}_i^* \\ \mathbf{Y}_i^{*'} \mathbf{y}_i & \mathbf{Y}_i^{*'} \mathbf{Y}_i & \mathbf{Y}_i^{*'} \mathbf{Y}_i^* \end{bmatrix} = \begin{bmatrix} \pi'_{.1} \mathbf{\Sigma}_{\mathbf{XX}} \pi_{.1} & \pi'_{.1} \mathbf{\Sigma}_{\mathbf{XX}} \pi_{.2} & \pi'_{.1} \mathbf{\Sigma}_{\mathbf{XX}} \pi_{.3} \\ \pi'_{.2} \mathbf{\Sigma}_{\mathbf{XX}} \pi_{.1} & \pi'_{.2} \mathbf{\Sigma}_{\mathbf{XX}} \pi_{.2} & \pi'_{.2} \mathbf{\Sigma}_{\mathbf{XX}} \pi_{.3} \\ \pi'_{.3} \mathbf{\Sigma}_{\mathbf{XX}} \pi_{.1} & \pi'_{.3} \mathbf{\Sigma}_{\mathbf{XX}} \pi_{.2} & \pi'_{.3} \mathbf{\Sigma}_{\mathbf{XX}} \pi_{.3} \end{bmatrix} + \mathbf{\Sigma}^v. \end{aligned}$$

Donde se retira que

$$p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{Y}'_i \mathbf{Y}_i) = \pi'_{.2} \mathbf{\Sigma}_{\mathbf{XX}} \pi_{.2} + \mathbf{\Sigma}^v_{22},$$

onde  $\mathbf{\Sigma}^v_{22}((M_i - 1) \times (M_i - 1))$  é a submatriz de  $\mathbf{\Sigma}^v$  correspondente a  $\mathbf{V}_i$ , de acordo com o desenvolvimento associado à partição considerada, que se apresenta de seguida,

$$\begin{aligned} p \lim_{T \rightarrow \infty} \left[ \frac{\mathbf{V}'\mathbf{V}}{T} \right] &= \mathbf{\Sigma}^v \Leftrightarrow p \lim_{T \rightarrow \infty} \frac{1}{T} \left\{ \begin{bmatrix} \mathbf{v}'_i \\ \mathbf{V}'_i \\ \mathbf{V}_i^{*'} \end{bmatrix} \begin{bmatrix} \mathbf{v}_i & \mathbf{V}_i & \mathbf{V}_i^* \end{bmatrix} \right\} = \mathbf{\Sigma}^v \\ &\Leftrightarrow p \lim_{T \rightarrow \infty} \frac{1}{T} \underbrace{\begin{bmatrix} \mathbf{v}'_i \mathbf{v}_i & \mathbf{v}'_i \mathbf{V}_i & \mathbf{v}'_i \mathbf{V}_i^* \\ \mathbf{V}'_i \mathbf{v}_i & \mathbf{V}'_i \mathbf{V}_i & \mathbf{V}'_i \mathbf{V}_i^* \\ \mathbf{V}_i^{*'} \mathbf{v}_i & \mathbf{V}_i^{*'} \mathbf{V}_i & \mathbf{V}_i^{*'} \mathbf{V}_i^* \end{bmatrix}}_{((1+(M_i - 1) + M_i^*) \times (1+(M_i - 1) + M_i^*))} = \mathbf{\Sigma}^v_{(M \times M)} \\ &\Rightarrow \mathbf{\Sigma}^v = \begin{bmatrix} \mathbf{\Sigma}^v_{11} & \mathbf{\Sigma}^v_{12} & \mathbf{\Sigma}^v_{13} \\ \mathbf{\Sigma}^v_{21} & \mathbf{\Sigma}^v_{22} & \mathbf{\Sigma}^v_{23} \\ \mathbf{\Sigma}^v_{31} & \mathbf{\Sigma}^v_{32} & \mathbf{\Sigma}^v_{33} \end{bmatrix}. \end{aligned}$$

Por outro lado, ainda pela **Proposição 2.2.3**, tem-se que

$$p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{Y}'\mathbf{X}) = \boldsymbol{\Pi}'\boldsymbol{\Sigma}_{\mathbf{XX}}.$$

Tendo em conta que

$$\begin{aligned} p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{Y}'\mathbf{X}) &= p \lim_{T \rightarrow \infty} \frac{1}{T} \left\{ \begin{bmatrix} \mathbf{y}'_i \\ \mathbf{Y}'_i \\ \mathbf{Y}_i^{*'} \end{bmatrix} \begin{bmatrix} \mathbf{X}_i & \mathbf{X}_i^* \end{bmatrix} \right\} \\ &= p \lim_{T \rightarrow \infty} \frac{1}{T} \begin{bmatrix} \mathbf{y}'_i\mathbf{X}_i & \mathbf{y}'_i\mathbf{X}_i^* \\ \mathbf{Y}'_i\mathbf{X}_i & \mathbf{Y}'_i\mathbf{X}_i^* \\ \mathbf{Y}_i^{*'}\mathbf{X}_i & \mathbf{Y}_i^{*'}\mathbf{X}_i^* \end{bmatrix}, \end{aligned}$$

e que

$$\begin{aligned} \boldsymbol{\Pi}'\boldsymbol{\Sigma}_{\mathbf{XX}} &= \begin{bmatrix} \pi'_{11} & \pi'_{21} \\ \pi'_{12} & \pi'_{22} \\ \pi'_{13} & \pi'_{23} \end{bmatrix} \begin{bmatrix} (\boldsymbol{\Sigma}_{\mathbf{XX}})_{11} & (\boldsymbol{\Sigma}_{\mathbf{XX}})_{12} \\ (\boldsymbol{\Sigma}_{\mathbf{XX}})_{21} & (\boldsymbol{\Sigma}_{\mathbf{XX}})_{22} \end{bmatrix} \\ &= \begin{bmatrix} \pi'_{.1}(\boldsymbol{\Sigma}_{\mathbf{XX}})_{.1} & \pi'_{.1}(\boldsymbol{\Sigma}_{\mathbf{XX}})_{.2} \\ \pi'_{.2}(\boldsymbol{\Sigma}_{\mathbf{XX}})_{.1} & \pi'_{.2}(\boldsymbol{\Sigma}_{\mathbf{XX}})_{.2} \\ \pi'_{.3}(\boldsymbol{\Sigma}_{\mathbf{XX}})_{.1} & \pi'_{.3}(\boldsymbol{\Sigma}_{\mathbf{XX}})_{.2} \end{bmatrix}, \end{aligned}$$

resulta que

$$p \lim_{T \rightarrow \infty} \frac{1}{T} \begin{bmatrix} \mathbf{y}'_i\mathbf{X}_i & \mathbf{y}'_i\mathbf{X}_i^* \\ \mathbf{Y}'_i\mathbf{X}_i & \mathbf{Y}'_i\mathbf{X}_i^* \\ \mathbf{Y}_i^{*'}\mathbf{X}_i & \mathbf{Y}_i^{*'}\mathbf{X}_i^* \end{bmatrix} = \begin{bmatrix} \pi'_{.1}(\boldsymbol{\Sigma}_{\mathbf{XX}})_{.1} & \pi'_{.1}(\boldsymbol{\Sigma}_{\mathbf{XX}})_{.2} \\ \pi'_{.2}(\boldsymbol{\Sigma}_{\mathbf{XX}})_{.1} & \pi'_{.2}(\boldsymbol{\Sigma}_{\mathbf{XX}})_{.2} \\ \pi'_{.3}(\boldsymbol{\Sigma}_{\mathbf{XX}})_{.1} & \pi'_{.3}(\boldsymbol{\Sigma}_{\mathbf{XX}})_{.2} \end{bmatrix},$$

e, consequentemente,

$$p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{Y}'_i\mathbf{X}_i) = \pi'_{.2}(\boldsymbol{\Sigma}_{\mathbf{XX}})_{.1}.$$

Por outro lado, como

$$\boldsymbol{\Sigma}_{\mathbf{XX}} = p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{X}'\mathbf{X}) \text{ e } \boldsymbol{\Sigma}_{\mathbf{XX}} = \begin{bmatrix} (\boldsymbol{\Sigma}_{\mathbf{XX}})_{11} & (\boldsymbol{\Sigma}_{\mathbf{XX}})_{12} \\ (\boldsymbol{\Sigma}_{\mathbf{XX}})_{21} & (\boldsymbol{\Sigma}_{\mathbf{XX}})_{22} \end{bmatrix},$$

atendendo a (3.3.18), vem que

$$p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{X}'_i\mathbf{X}_i) = (\boldsymbol{\Sigma}_{\mathbf{XX}})_{11}.$$



Por fim, como

$$\begin{aligned}
 p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{Y}' \mathbf{E}) &= p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{\Pi}' \mathbf{X}' + \mathbf{V}') \mathbf{E} \\
 &= p \lim_{T \rightarrow \infty} \frac{1}{T} \mathbf{\Pi}' \mathbf{X}' \mathbf{E} + p \lim_{T \rightarrow \infty} \frac{1}{T} \mathbf{V}' \mathbf{E} \\
 &= \mathbf{\Pi}' \mathbf{0} + p \lim_{T \rightarrow \infty} \frac{1}{T} \left[ (\mathbf{\Gamma}^{-1})' \mathbf{E}' \mathbf{E} \right], \text{ por (2.2.13)} \\
 &= (\mathbf{\Gamma}^{-1})' \mathbf{\Sigma}, \text{ por HFE.5,}
 \end{aligned}$$

e considerando que

$$\begin{aligned}
 p \lim_{T \rightarrow \infty} \frac{1}{T} \mathbf{Y}' \mathbf{E} &= p \lim_{T \rightarrow \infty} \frac{1}{T} \left\{ \begin{bmatrix} \mathbf{y}'_i \\ \mathbf{Y}'_i \\ \mathbf{Y}_i^{*'} \end{bmatrix} \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_i & \cdots & \mathbf{e}_M \end{bmatrix} \right\} \\
 &= p \lim_{T \rightarrow \infty} \frac{1}{T} \begin{bmatrix} \mathbf{y}'_i \mathbf{e}_1 & \cdots & \mathbf{y}'_i \mathbf{e}_i & \cdots & \mathbf{y}'_i \mathbf{e}_M \\ \mathbf{Y}'_i \mathbf{e}_1 & \cdots & \mathbf{Y}'_i \mathbf{e}_i & \cdots & \mathbf{Y}'_i \mathbf{e}_M \\ \mathbf{Y}_i^{*'} \mathbf{e}_1 & \cdots & \mathbf{Y}_i^{*'} \mathbf{e}_i & \cdots & \mathbf{Y}_i^{*'} \mathbf{e}_M \end{bmatrix},
 \end{aligned}$$

vem

$$p \lim_{T \rightarrow \infty} \frac{1}{T} (\mathbf{Y}'_i \mathbf{e}_i) = \left[ (\mathbf{\Gamma}^{-1})' \right]_* \mathbf{\Sigma}_{.i},$$

onde  $\left[ (\mathbf{\Gamma}^{-1})' \right]_* \mathbf{\Sigma}_{.i}$  representa a submatriz de  $(\mathbf{\Gamma}^{-1})' \mathbf{\Sigma}$  correspondente ao limite do bloco  $\mathbf{Y}'_i \mathbf{e}_i$ ,  $\mathbf{\Sigma}_{.i}$  é a coluna  $i$  da matriz  $\mathbf{\Sigma}$  e  $\left[ (\mathbf{\Gamma}^{-1})' \right]_*$  é a submatriz de  $(\mathbf{\Gamma}^{-1})'$  cujos elementos correspondem a  $\mathbf{Y}_i$ .

Além disso, como  $p \lim_{T \rightarrow \infty} \frac{1}{T} \mathbf{X}' \mathbf{E} = \mathbf{0}$ , então  $p \lim_{T \rightarrow \infty} \frac{1}{T} \mathbf{X}'_i \mathbf{e}_i = \mathbf{0}$ .

Conjugando estes resultados, resulta que partindo de (7.2.2) se tem

$$\begin{aligned}
 p \lim_{T \rightarrow \infty} \left[ \hat{\boldsymbol{\delta}}_{i,OLS} \right] &= \boldsymbol{\delta}_i + p \lim_{T \rightarrow \infty} \begin{bmatrix} \frac{\mathbf{Y}'_i \mathbf{Y}_i}{T} & \frac{\mathbf{Y}'_i \mathbf{X}_i}{T} \\ \frac{\mathbf{X}'_i \mathbf{Y}_i}{T} & \frac{\mathbf{X}'_i \mathbf{X}_i}{T} \end{bmatrix}^{-1} \times p \lim_{T \rightarrow \infty} \begin{bmatrix} \frac{\mathbf{Y}'_i \mathbf{e}_i}{T} \\ \frac{\mathbf{X}'_i \mathbf{e}_i}{T} \end{bmatrix} \\
 &= \boldsymbol{\delta}_i + p \lim_{T \rightarrow \infty} \begin{bmatrix} \boldsymbol{\pi}'_{.2} \mathbf{\Sigma}_{\mathbf{XX}} \boldsymbol{\pi}_{.2} + \Sigma^v_{22} & \boldsymbol{\pi}'_{.2} (\mathbf{\Sigma}_{\mathbf{XX}})_{.1} \\ (\mathbf{\Sigma}_{\mathbf{XX}})'_{.1} \boldsymbol{\pi}_{.2} & (\mathbf{\Sigma}_{\mathbf{XX}})_{11} \end{bmatrix}^{-1} \times \begin{bmatrix} \left[ (\mathbf{\Gamma}^{-1})' \right]_* \mathbf{\Sigma}_{.i} \\ \mathbf{0} \end{bmatrix} \\
 &\neq \boldsymbol{\delta}_i,
 \end{aligned}$$

o que traduz que o estimador  $\hat{\boldsymbol{\delta}}_{i,OLS}$  não é consistente para  $\boldsymbol{\delta}_i$ . □

A.2.2 O estimador de variáveis instrumentais goza das seguintes propriedades:

PVI.1 O estimador de variáveis instrumentais,  $\hat{\boldsymbol{\delta}}_{i,VI}$ , é um estimador consistente para  $\boldsymbol{\delta}_i$ .

*Demonstração.* Para verificar esta propriedade, é preciso averiguar se  $p \lim_{T \rightarrow \infty} [\hat{\boldsymbol{\delta}}_{i,VI}] = \boldsymbol{\delta}_i$ .

$$\begin{aligned}
 p \lim_{T \rightarrow \infty} [\hat{\boldsymbol{\delta}}_{i,VI}] &= p \lim_{T \rightarrow \infty} [(\mathbf{W}'_i \mathbf{Z}_i)^{-1} \mathbf{W}'_i \mathbf{y}_i] \\
 &= p \lim_{T \rightarrow \infty} [(\mathbf{W}'_i \mathbf{Z}_i)^{-1} \mathbf{W}'_i (\mathbf{Z}_i \boldsymbol{\delta}_i + \mathbf{e}_i)] \\
 &= p \lim_{T \rightarrow \infty} [(\mathbf{W}'_i \mathbf{Z}_i)^{-1} \mathbf{W}'_i \mathbf{Z}_i \boldsymbol{\delta}_i + (\mathbf{W}'_i \mathbf{Z}_i)^{-1} \mathbf{W}'_i \mathbf{e}_i] \\
 &= \boldsymbol{\delta}_i + p \lim_{T \rightarrow \infty} [(\mathbf{W}'_i \mathbf{Z}_i)^{-1} \mathbf{W}'_i \mathbf{e}_i] \\
 &= \boldsymbol{\delta}_i + p \lim_{T \rightarrow \infty} \left( \frac{\mathbf{W}'_i \mathbf{Z}_i}{T} \right)^{-1} p \lim_{T \rightarrow \infty} \frac{\mathbf{W}'_i \mathbf{e}_i}{T} \\
 &= \boldsymbol{\delta}_i + \boldsymbol{\Sigma}_{W_i Z_i}^{-1} \times \mathbf{0}, \text{ por (3.3.9) e (3.3.10)} \\
 &= \boldsymbol{\delta}_i.
 \end{aligned} \tag{7.2.6}$$

□

IPVI.2 Se  $\frac{\mathbf{W}'_i \mathbf{e}_i}{\sqrt{T}}$  converge em distribuição para a  $N(\mathbf{0}, \sigma_i^2 \boldsymbol{\Sigma}_{W_i W_i})$ , então

$$\sqrt{T} (\hat{\boldsymbol{\delta}}_{i,VI} - \boldsymbol{\delta}_i) \stackrel{\circ}{\sim} N(\mathbf{0}, \sigma_i^2 [\boldsymbol{\Sigma}_{W_i Z_i}^{-1} \boldsymbol{\Sigma}_{W_i W_i} \boldsymbol{\Sigma}_{Z_i W_i}^{-1}]).$$

*Demonstração.* Tendo em conta (7.2.6), vem

$$\sqrt{T} (\hat{\boldsymbol{\delta}}_{i,VI} - \boldsymbol{\delta}_i) = \left( \frac{\mathbf{W}'_i \mathbf{Z}_i}{T} \right)^{-1} \frac{\mathbf{W}'_i \mathbf{e}_i}{\sqrt{T}}.$$

Como  $p \lim_{T \rightarrow \infty} \left[ \frac{\mathbf{W}'_i \mathbf{Z}_i}{T} \right]^{-1} = \boldsymbol{\Sigma}_{W_i Z_i}^{-1}$  existe e é finita, de acordo com (3.3.8), então  $\sqrt{T} (\hat{\boldsymbol{\delta}}_{i,VI} - \boldsymbol{\delta}_i)$  terá a mesma distribuição limite que  $\frac{\mathbf{W}'_i \mathbf{e}_i}{\sqrt{T}}$ . Se, por hipótese,  $\frac{\mathbf{W}'_i \mathbf{e}_i}{\sqrt{T}} \stackrel{\circ}{\sim} N(\mathbf{0}, \sigma_i^2 \boldsymbol{\Sigma}_{W_i W_i})$ , então

$$\sqrt{T} (\hat{\boldsymbol{\delta}}_{i,VI} - \boldsymbol{\delta}_i) \stackrel{\circ}{\sim} N(\mathbf{0}, \sigma_i^2 [\boldsymbol{\Sigma}_{W_i Z_i}^{-1} \boldsymbol{\Sigma}_{W_i W_i} (\boldsymbol{\Sigma}_{W_i Z_i}^{-1})']).$$

□

IPVI.3 A matriz de covariâncias assintótica de  $\hat{\boldsymbol{\delta}}_{i,VI}$  é

$$VA [\hat{\boldsymbol{\delta}}_{i,VI}] = \frac{\sigma_i^2}{T} [\boldsymbol{\Sigma}_{W_i Z_i}^{-1} \boldsymbol{\Sigma}_{W_i W_i} \boldsymbol{\Sigma}_{Z_i W_i}^{-1}],$$

e um estimador consistente para  $\sigma_i^2$  é  $\hat{\sigma}_i^2 = \frac{1}{T} \left[ (\mathbf{y}_i - \mathbf{Z}_i \hat{\boldsymbol{\delta}}_{i,VI})' (\mathbf{y}_i - \mathbf{Z}_i \hat{\boldsymbol{\delta}}_{i,VI}) \right]$ .

*Demonstração.* Para estimar  $\sigma_i^2$ , a variância do erro aleatório da equação  $i$ ,  $\mathbf{e}_i$ , e a variância dos resíduos relativos à equação  $i$  aparece como um estimador natural,

$$\begin{aligned}\hat{\sigma}_i^2 &= \frac{1}{T} [\hat{\mathbf{e}}_i' \hat{\mathbf{e}}_i] \\ &= \frac{1}{T} \left[ \left( \mathbf{y}_i - \mathbf{Z}_i \hat{\delta}_{i,VI} \right)' \left( \mathbf{y}_i - \mathbf{Z}_i \hat{\delta}_{i,VI} \right) \right] \\ &= \frac{1}{T} \sum_{t=1}^T \left( \mathbf{y}_{ti} - \mathbf{Z}_{ti} \hat{\delta}_{i,VI} \right)^2.\end{aligned}$$

Partindo do vector de resíduos da equação  $i$ ,

$$\begin{aligned}\hat{\mathbf{e}}_i &= \mathbf{y}_i - \mathbf{Z}_i \hat{\delta}_{i,VI} \\ &= \mathbf{y}_i - \mathbf{Z}_i (\mathbf{W}_i' \mathbf{Z}_i)^{-1} \mathbf{W}_i' \mathbf{y}_i,\end{aligned}$$

e substituindo  $\mathbf{y}_i = \mathbf{Z}_i \boldsymbol{\delta}_i + \mathbf{e}_i$ , vem

$$\begin{aligned}\mathbf{y}_i - \mathbf{Z}_i \hat{\delta}_{i,VI} &= (\mathbf{Z}_i \boldsymbol{\delta}_i + \mathbf{e}_i) - \mathbf{Z}_i (\mathbf{W}_i' \mathbf{Z}_i)^{-1} \mathbf{W}_i' (\mathbf{Z}_i \boldsymbol{\delta}_i + \mathbf{e}_i) \\ &= (\mathbf{Z}_i \boldsymbol{\delta}_i + \mathbf{e}_i) - \mathbf{Z}_i (\mathbf{W}_i' \mathbf{Z}_i)^{-1} \mathbf{W}_i' \mathbf{Z}_i \boldsymbol{\delta}_i - \mathbf{Z}_i (\mathbf{W}_i' \mathbf{Z}_i)^{-1} \mathbf{W}_i' \mathbf{e}_i \\ &= (\mathbf{Z}_i \boldsymbol{\delta}_i + \mathbf{e}_i) - \mathbf{Z}_i \boldsymbol{\delta}_i - \mathbf{Z}_i (\mathbf{W}_i' \mathbf{Z}_i)^{-1} \mathbf{W}_i' \mathbf{e}_i \\ &= \mathbf{e}_i - \mathbf{Z}_i (\mathbf{W}_i' \mathbf{Z}_i)^{-1} \mathbf{W}_i' \mathbf{e}_i \\ &= \mathbf{e}_i \left[ \mathbf{I} - \mathbf{Z}_i (\mathbf{W}_i' \mathbf{Z}_i)^{-1} \mathbf{W}_i' \right] \\ &= \hat{\mathbf{e}}_i.\end{aligned}$$

Deste modo,

$$\begin{aligned}\hat{\sigma}_i^2 &= \frac{1}{T} (\hat{\mathbf{e}}_i' \hat{\mathbf{e}}_i) \\ &= \frac{1}{T} \left[ \mathbf{e}_i \left( \mathbf{I} - \mathbf{Z}_i (\mathbf{W}_i' \mathbf{Z}_i)^{-1} \mathbf{W}_i' \right) \right]' \left[ \mathbf{e}_i \left( \mathbf{I} - \mathbf{Z}_i (\mathbf{W}_i' \mathbf{Z}_i)^{-1} \mathbf{W}_i' \right) \right] \\ &= \frac{1}{T} \left[ \mathbf{e}_i - \mathbf{Z}_i (\mathbf{W}_i' \mathbf{Z}_i)^{-1} \mathbf{W}_i' \mathbf{e}_i \right]' \left[ \mathbf{e}_i - \mathbf{Z}_i (\mathbf{W}_i' \mathbf{Z}_i)^{-1} \mathbf{W}_i' \mathbf{e}_i \right] \\ &= \frac{1}{T} \left[ \mathbf{e}_i' - \mathbf{e}_i' \mathbf{W}_i (\mathbf{Z}_i' \mathbf{W}_i)^{-1} \mathbf{Z}_i' \right] \left[ \mathbf{e}_i - \mathbf{Z}_i (\mathbf{W}_i' \mathbf{Z}_i)^{-1} \mathbf{W}_i' \mathbf{e}_i \right] \\ &= \frac{\mathbf{e}_i' \mathbf{e}_i}{T} - 2 \left( \frac{\mathbf{e}_i' \mathbf{W}_i}{T} \right) \left( \frac{\mathbf{Z}_i' \mathbf{W}_i}{T} \right)^{-1} \left( \frac{\mathbf{Z}_i' \mathbf{e}_i}{T} \right) \\ &\quad + \left( \frac{\mathbf{e}_i' \mathbf{W}_i}{T} \right) \left( \frac{\mathbf{Z}_i' \mathbf{W}_i}{T} \right)^{-1} \left( \frac{\mathbf{Z}_i' \mathbf{Z}_i}{T} \right) \left( \frac{\mathbf{W}_i' \mathbf{Z}_i}{T} \right)^{-1} \left( \frac{\mathbf{W}_i' \mathbf{e}_i}{T} \right).\end{aligned}$$

Calculando o limite em probabilidade de  $\hat{\sigma}_i^2$ , obtém-se apenas  $\sigma_i^2$ , devido à primeira parcela, pois a segunda e a terceira parcelas convergem para zero, atendendo a (3.3.8) e (3.3.9). Conclui-se assim que  $\hat{\sigma}_i^2$  é um estimador consistente para  $\sigma_i^2$ .  $\square$

### A.2.3 Demonstração do Teorema 3.9.5.

*Demonstração.* Para comparar as matrizes de covariâncias assintóticas dos estimadores, partiu-se das expressões (3.9.2) e (3.9.1), ou seja,

$$VA(\hat{\delta}_{GMMIL}) = p \lim [\mathbf{C}'(\boldsymbol{\Omega}_{IL}^*)^{-1} \mathbf{C}]^{-1} \text{ e } VA(\hat{\delta}_{GMMIC}) = p \lim [\mathbf{C}'(\boldsymbol{\Omega}^*)^{-1} \mathbf{C}]^{-1},$$

e observou-se que as matrizes têm uma forma idêntica, diferindo apenas nas matrizes de covariâncias dos erros do modelo,  $\boldsymbol{\Omega}_{IL}$  e  $\boldsymbol{\Omega}$ . Deste modo, calcularam-se e compararam-se os determinantes das matrizes  $\boldsymbol{\Omega}_{IL}$  e  $\boldsymbol{\Omega}$ .

Como são matrizes definidas por blocos, usou-se uma decomposição por blocos sugerida em Seber (1984). O estudo foi desenvolvido em duas partes, primeiro tratou-se o caso de um modelo com duas equações ( $M = 2$ ), calcularam-se os determinantes, e estabeleceu-se a sua comparação. De seguida procedeu-se à generalização para o caso de um modelo com  $M$  equações, por indução.

No caso de um modelo formado por duas equações ( $M = 2$ ), temos as matrizes  $\boldsymbol{\Omega}_{IL}$  e  $\boldsymbol{\Omega}$ , com dimensão  $(2T \times 2T)$ , da seguinte forma

$$\boldsymbol{\Omega}_{IL} = \begin{bmatrix} \boldsymbol{\Omega}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_{22} \end{bmatrix} (MIL),$$

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{bmatrix} (MIC).$$

No cálculo do determinante,  $|\boldsymbol{\Omega}|$ , teve-se em conta que

$$\begin{bmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\boldsymbol{\Omega}_{11}^{-1} \boldsymbol{\Omega}_{12} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Omega}_{11} & \mathbf{0} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{21} \boldsymbol{\Omega}_{11}^{-1} \boldsymbol{\Omega}_{12} \end{bmatrix}. \quad (7.2.7)$$

Calculando os determinantes de ambos os membros de (7.2.7) e pelas propriedades dos determinantes, vem que

$$|\boldsymbol{\Omega}| = |\boldsymbol{\Omega}_{11}| |\boldsymbol{\Omega}_{22 \cdot 1}|,$$

onde

$$\boldsymbol{\Omega}_{22 \cdot 1} = \boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{21} \boldsymbol{\Omega}_{11}^{-1} \boldsymbol{\Omega}_{12}. \quad (7.2.8)$$

A matriz  $\boldsymbol{\Omega}_{22 \cdot 1}$  representa a matriz de covariâncias de  $\mathbf{e}_2 | \mathbf{e}_1$ , ou a matriz de covariâncias parciais de  $\mathbf{e}_2$  dado  $\mathbf{e}_1$ , de acordo com Mardia (1979).

De seguida, para obter a forma de  $\mathbf{\Omega}_{22.1}$ , considerou-se que

$$\mathbf{\Omega}_{22} = \begin{bmatrix} \sigma_{21}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{2T}^2 \end{bmatrix}, \mathbf{\Omega}_{21} = \begin{bmatrix} \sigma_{211} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{21T} \end{bmatrix}, \mathbf{\Omega}_{11}^{-1} = \begin{bmatrix} \sigma_{11}^{-2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{1T}^{-2} \end{bmatrix}.$$

Efectuaram-se alguns cálculos, resultando que

$$\begin{aligned} \mathbf{\Omega}_{22.1} &= \text{diag} \left[ \sigma_{2,t}^2 - \frac{\sigma_{21t}\sigma_{12t}}{\sigma_{1t}^2} \right] \\ &= \text{diag} \left[ \sigma_{2t}^2 \left( 1 - \frac{\sigma_{12t}^2}{\sigma_{1,t}^2 \sigma_{2,t}^2} \right) \right] \\ &= \text{diag} \left[ \sigma_{2t}^2 (1 - \rho_{12t}^2) \right]_{t=1, \dots, T}, \end{aligned}$$

onde  $\rho_{12t} = \frac{\sigma_{12t}}{\sigma_{1t}\sigma_{2t}}$  representa o coeficiente de correlação entre  $e_{1t}$  e  $e_{2t}$ , e, de um modo mais geral,

$$\rho_{ij,t} = \frac{\sigma_{ijt}}{\sigma_{it}^2 \sigma_{jt}^2},$$

representa o coeficiente de correlação entre  $e_{it}$  e  $e_{jt}$ .

Por outro lado, note-se que  $\sigma_{21,t} = E[e_{2t}e_{1t}] = E[e_{1t}e_{2t}] = \sigma_{12,t}, \forall t = 1, \dots, T$ , e também  $\sigma_{ijt} = \sigma_{ji,t}$ . Deste modo vem que

$$|\mathbf{\Omega}| = |\mathbf{\Omega}_{11}| |\mathbf{\Omega}_{22.1}| \quad (7.2.9)$$

$$\begin{aligned} &= |\text{diag} [\sigma_{1t}^2]| |\text{diag} [\sigma_{2t}^2 (1 - \rho_{12t}^2)]| \\ &= |\text{diag} [\sigma_{1t}^2 \sigma_{2t}^2 (1 - \rho_{12t}^2)]| \\ &= |\mathbf{\Omega}_{11}| |\mathbf{\Omega}_{22}| |\text{diag} [(1 - \rho_{12t}^2)]| \\ &= |\mathbf{\Omega}_{11}| |\mathbf{\Omega}_{22}| \prod_{t=1}^T (1 - \rho_{12t}^2). \end{aligned} \quad (7.2.10)$$

Tendo em conta que

$$\begin{aligned} 0 \leq \rho_{ijt}^2 \leq 1 &\Leftrightarrow 0 \leq 1 - \rho_{ijt}^2 \leq 1 \\ \Rightarrow 0 \leq \prod_{t=1}^T (1 - \rho_{ijt}^2) &\leq 1, \end{aligned}$$

por (7.2.10), vem que

$$|\mathbf{\Omega}| \leq \underbrace{|\mathbf{\Omega}_{11}| |\mathbf{\Omega}_{22}|}_{|\mathbf{\Omega}_{LL}|}.$$

Fica assim provado que, no caso em que  $M = 2$ , se tem

$$|\Omega| \leq |\Omega_{IL}|.$$

Para estabelecer as expressões das matrizes e dos determinantes correspondentes ao caso de um sistema com  $M$  equações, tornou-se necessário obter previamente essas expressões nos casos em que  $M = 3$  e  $M = 4$ . No caso de um modelo com três equações,  $M = 3$ , as matrizes  $\Omega_{IL}$  e  $\Omega$ , com dimensão  $(3T \times 3T)$ , têm a seguinte forma:

$$\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} \\ \Omega_{21} & \Omega_{22} & \Omega_{23} \\ \Omega_{31} & \Omega_{32} & \Omega_{33} \end{bmatrix} \text{ e } \Omega_{IL} = \begin{bmatrix} \Omega_{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Omega_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Omega_{33} \end{bmatrix}.$$

Seguindo um procedimento análogo ao utilizado no caso anterior em (7.2.9), determinou-se uma matriz auxiliar,  $\mathfrak{A}$ , triangular superior, da forma  $\mathfrak{A} = \begin{bmatrix} \mathbf{I} & \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{I} & \mathbf{C} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}$ , de modo

que o produto  $\Omega\mathfrak{A}$  fosse uma matriz triangular inferior, ou seja,

$$\Omega\mathfrak{A} = \begin{bmatrix} \mathbf{D} & \mathbf{0} & \mathbf{0} \\ \mathbf{E} & \mathbf{F} & \mathbf{0} \\ \mathbf{G} & \mathbf{H} & \mathbf{J} \end{bmatrix}.$$

Para determinar a matriz  $\mathfrak{A}$ , foi necessário calcular as matrizes  $\mathbf{A}$ ,  $\mathbf{B}$  e  $\mathbf{C}$ , que são os blocos desconhecidos dessa matriz. Para isso, resolveu-se a equação matricial

$$\begin{aligned} \Omega\mathfrak{A} &= \begin{bmatrix} \mathbf{D} & \mathbf{0} & \mathbf{0} \\ \mathbf{E} & \mathbf{F} & \mathbf{0} \\ \mathbf{G} & \mathbf{H} & \mathbf{J} \end{bmatrix} \\ \Leftrightarrow \begin{bmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} \\ \Omega_{21} & \Omega_{22} & \Omega_{23} \\ \Omega_{31} & \Omega_{32} & \Omega_{33} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{I} & \mathbf{C} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} &= \begin{bmatrix} \mathbf{D} & \mathbf{0} & \mathbf{0} \\ \mathbf{E} & \mathbf{F} & \mathbf{0} \\ \mathbf{G} & \mathbf{H} & \mathbf{J} \end{bmatrix} \\ \Leftrightarrow \begin{cases} \Omega_{11}\mathbf{A} + \Omega_{12} = \mathbf{0} \\ \Omega_{11}\mathbf{B} + \Omega_{12}\mathbf{C} + \Omega_{13} = \mathbf{0} \\ \Omega_{21}\mathbf{B} + \Omega_{22}\mathbf{C} + \Omega_{23} = \mathbf{0} \end{cases} & \quad (i) \\ \Leftrightarrow \begin{cases} \mathbf{A} = -\Omega_{11}^{-1}\Omega_{12} \\ \mathbf{B} = \Omega_{11}^{-1} \left( \Omega_{12} (\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1} (\Omega_{23} - \Omega_{21}\Omega_{11}^{-1}\Omega_{13}) - \Omega_{13} \right) \\ \mathbf{C} = -(\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1} (\Omega_{23} - \Omega_{21}\Omega_{11}^{-1}\Omega_{13}) \end{cases} \end{aligned}$$

Deste modo, vem  $\mathfrak{A} = \begin{bmatrix} \mathbf{I} & \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{I} & \mathbf{C} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}$ , onde os blocos  $\mathbf{A}, \mathbf{B}$  e  $\mathbf{C}$  são as soluções do sistema anterior. Uma vez que  $|\mathfrak{A}| = 1$ , vem que  $|\Omega\mathfrak{A}| = |\Omega| |\mathfrak{A}| = |\Omega|$ , logo

$$|\Omega| = |\Omega\mathfrak{A}| = |\Omega_{11}| \times |(\Omega_{21}\mathbf{A} + \Omega_{22}) \times (\Omega_{31}\mathbf{B} + \Omega_{32}\mathbf{C} + \Omega_{33})|.$$

Calculou-se primeiro o produto das matrizes

$$\begin{aligned} & (\Omega_{21}\mathbf{A} + \Omega_{22}) \times (\Omega_{31}\mathbf{B} + \Omega_{32}\mathbf{C} + \Omega_{33}) = (\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}) \\ & \times [-\Omega_{31}\Omega_{11}^{-1}(\Omega_{13} - \Omega_{12}(\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1}(\Omega_{23} - \Omega_{21}\Omega_{11}^{-1}\Omega_{13}) \\ & - \Omega_{32}(\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1}(\Omega_{23} - \Omega_{21}\Omega_{11}^{-1}\Omega_{13}) + \Omega_{33}] \\ & = \Omega_{22 \cdot 1} [(\Omega_{33} - \Omega_{31}\Omega_{11}^{-1}\Omega_{13}) - (\Omega_{32} - \Omega_{31}\Omega_{11}^{-1}\Omega_{12})\Omega_{22 \cdot 1}^{-1}\Omega_{23 \cdot 1}] \\ & = \Omega_{22 \cdot 1} [\Omega_{33 \cdot 1} - \Omega_{32 \cdot 1}\Omega_{22 \cdot 1}^{-1}\Omega_{23 \cdot 1}] = \Omega_{22 \cdot 1}\Omega_{33 \cdot 21}. \end{aligned}$$

Note-se que, à semelhança de (7.2.8),

$$\Omega_{23 \cdot 1} = \Omega_{23} - \Omega_{21}\Omega_{11}^{-1}\Omega_{31}$$

representa a matriz de covariâncias parciais entre  $\mathbf{e}_2$  e  $\mathbf{e}_3$  dado  $\mathbf{e}_1$ , e que

$$\Omega_{33 \cdot 21} = \Omega_{33 \cdot 1} - \Omega_{32 \cdot 1}\Omega_{22 \cdot 1}^{-1}\Omega_{23 \cdot 1}$$

representa a matriz de covariâncias parciais de  $\mathbf{e}_3$  dado  $\mathbf{e}_1$  e  $\mathbf{e}_2$ .

Calculou-se depois o determinante de  $\Omega$ , considerando

$$|\Omega| = |\Omega_{11}| |\Omega_{22 \cdot 1}\Omega_{33 \cdot 21}|. \quad (7.2.11)$$

Usando o resultado obtido para o modelo com duas equações, tem-se por (7.2.10) que

$$|\Omega_{11}| |\Omega_{22 \cdot 1}| = |\Omega_{11}| |\Omega_{22}| \prod_{t=1}^T (1 - \rho_{12,t}^2),$$

pelo que falta apenas obter  $|\Omega_{33 \cdot 21}|$ . Para isso, teve-se em conta que

$$\Omega_{33 \cdot 21} = \Omega_{33 \cdot 2} - \Omega_{31 \cdot 2}\Omega_{11 \cdot 2}^{-1}\Omega_{13 \cdot 2}, \quad (7.2.12)$$

onde

$$\Omega_{33 \cdot 2} = \text{diag} [\sigma_{3t}^2 (1 - \rho_{32t}^2)],$$

$$\mathbf{\Omega}_{11 \cdot 2}^{-1} = \text{diag} \left[ \frac{1}{\sigma_{1t}^2 (1 - \rho_{12,t}^2)} \right],$$

$$\begin{aligned} \mathbf{\Omega}_{31 \cdot 2} &= \mathbf{\Omega}_{31} - \mathbf{\Omega}_{32} \mathbf{\Omega}_{22}^{-1} \mathbf{\Omega}_{12} \\ &= \text{diag} [\sigma_{31t}] - \text{diag} \left[ \frac{\sigma_{32t} \sigma_{21t}}{\sigma_{2t}^2} \right] \\ &= \text{diag} \left[ \sigma_{31t} - \frac{\sigma_{32t} \sigma_{21t}}{\sigma_{2t}^2} \right]_{t=1, \dots, T} \\ &= \mathbf{\Omega}_{13 \cdot 2}. \end{aligned}$$

Então vem

$$\begin{aligned} \mathbf{\Omega}_{33 \cdot 21} &= \text{diag} [\sigma_{3t}^2 (1 - \rho_{32t}^2)] \\ &\quad - \text{diag} \left[ \left( \sigma_{31t} - \frac{\sigma_{32t} \sigma_{21t}}{\sigma_{2t}^2} \right) \left( \frac{1}{\sigma_{1t}^2 (1 - \rho_{12t}^2)} \right) \left( \sigma_{31t} - \frac{\sigma_{32t} \sigma_{21t}}{\sigma_{2t}^2} \right) \right] \\ &= \text{diag} \left[ \sigma_{3t}^2 (1 - \rho_{32t}^2) - \frac{\left( \sigma_{31t} - \frac{\sigma_{32t} \sigma_{21t}}{\sigma_{2t}^2} \right)^2}{\sigma_{1t}^2 (1 - \rho_{12t}^2)} \right]_{t=1, \dots, T}. \end{aligned}$$

Tomando o termo de ordem  $t$  e tendo em conta a definição de coeficiente de correlação, vem que,

$$\begin{aligned} \sigma_{3t}^2 (1 - \rho_{32t}^2) - \frac{\left( \sigma_{31t} - \frac{\sigma_{32t} \sigma_{21t}}{\sigma_{2t}^2} \right)^2}{\sigma_{1t}^2 (1 - \rho_{12t}^2)} &= \\ &= \sigma_{3t}^2 (1 - \rho_{32t}^2) - \frac{\left( \frac{\sigma_{31t} \sigma_{2t}^2 - \sigma_{32t} \sigma_{21t}}{\sigma_{2t}^2} \right)^2}{\sigma_{1t}^2 (1 - \rho_{12t}^2)} \\ &= \frac{\sigma_{3t}^2 (1 - \rho_{32t}^2) \sigma_{2t}^4 \sigma_{1t}^2 (1 - \rho_{12t}^2) - (\sigma_{31t} \sigma_{2t}^2 - \sigma_{32t} \sigma_{21t})^2}{\sigma_{2t}^4 \sigma_{1t}^2 (1 - \rho_{12t}^2)} \\ &= \frac{\sigma_{3t}^2 \left[ (1 - \rho_{32t}^2) (1 - \rho_{12t}^2) - \left( \frac{\sigma_{31t}}{\sigma_{3t} \sigma_{1t}} - \frac{\sigma_{32t} \sigma_{21t}}{\sigma_{3t} \sigma_{1t} \sigma_{2t}^2} \right)^2 \right]}{(1 - \rho_{12t}^2)} \\ &= \frac{\sigma_{3t}^2}{(1 - \rho_{12t}^2)} \left[ (1 - \rho_{32t}^2) (1 - \rho_{12t}^2) - (\rho_{31t} - \rho_{32t} \rho_{12t})^2 \right]. \end{aligned}$$



Prosseguindo com o mesmo tipo de simplificações, vem

$$\begin{aligned}
& \frac{\sigma_{3t}^2}{(1 - \rho_{12t}^2)} \left[ (1 - \rho_{32t}^2) (1 - \rho_{12t}^2) - \frac{(\rho_{31t} - \rho_{32t}\rho_{12t})^2}{(1 - \rho_{32t}^2)(1 - \rho_{12t}^2)} (1 - \rho_{32t}^2) (1 - \rho_{12t}^2) \right] \\
&= \frac{\sigma_{3t}^2}{(1 - \rho_{12t}^2)} [(1 - \rho_{32t}^2) (1 - \rho_{12t}^2) - \rho_{31.2t}^2 (1 - \rho_{32t}^2) (1 - \rho_{12t}^2)] \quad (7.2.13) \\
&= \frac{\sigma_{3t}^2}{(1 - \rho_{12t}^2)} (1 - \rho_{32t}^2) (1 - \rho_{12t}^2) (1 - \rho_{31.2t}^2) = \sigma_{3t}^2 (1 - \rho_{32t}^2) (1 - \rho_{31.2t}^2).
\end{aligned}$$

Em (7.2.13), representou-se por

$$\rho_{31.2t}^2 = \frac{(\rho_{31t} - \rho_{32t}\rho_{12t})^2}{(1 - \rho_{32t}^2)(1 - \rho_{12t}^2)},$$

o coeficiente de correlação parcial entre  $e_{3t}$  e  $e_{1t}$  dado  $e_{2t}$ , e, no caso geral, usou-se

$$\rho_{ij.kt}^2 = \frac{(\rho_{ijt} - \rho_{ikt}\rho_{jkt})^2}{(1 - \rho_{ikt}^2)(1 - \rho_{jkt}^2)}, \quad (7.2.14)$$

para representar o coeficiente de correlação parcial entre  $e_{it}$  e  $e_{jt}$  dado  $e_{kt}$ .

Deste modo, vem que

$$\mathbf{\Omega}_{33.21} = \text{diag} [\sigma_{3t}^2 (1 - \rho_{32t}^2) (1 - \rho_{31.2t}^2)]_{t=1, \dots, T}. \quad (7.2.15)$$

Consequentemente,

$$\begin{aligned}
|\mathbf{\Omega}| &= |\mathbf{\Omega}_{11}| |\mathbf{\Omega}_{22}| \prod_{t=1}^T (1 - \rho_{12t}^2) |\text{diag} [\sigma_{3t}^2 (1 - \rho_{32t}^2) (1 - \rho_{31.2t}^2)]| \\
&= |\mathbf{\Omega}_{11}| |\mathbf{\Omega}_{22}| |\mathbf{\Omega}_{33}| \prod_{t=1}^T [(1 - \rho_{12t}^2) (1 - \rho_{32t}^2) (1 - \rho_{31.2t}^2)].
\end{aligned}$$

Do mesmo modo que para o coeficiente de correlação se tem  $0 \leq \rho^2 \leq 1$ , também para os coeficientes de correlação parciais se verifica que os seus quadrados tomam valores em  $[0, 1]$ , isto é,

$$0 \leq \rho_{31.2t}^2 \leq 1, 0 \leq \rho_{41.32t}^2 \leq 1, \dots$$

Assim, fazendo

$$a_t = (1 - \rho_{12t}^2) (1 - \rho_{32t}^2) (1 - \rho_{31.2t}^2),$$

resulta que  $0 \leq a_t \leq 1$ , uma vez que é o produto de vários factores que tomam valores em  $[0, 1]$ , e também que  $0 \leq \prod_{t=1}^T a_t \leq 1$ . Logo, pode concluir-se que

$$|\mathbf{\Omega}| \leq \underbrace{|\mathbf{\Omega}_{11}| |\mathbf{\Omega}_{22}| |\mathbf{\Omega}_{33}|}_{|\mathbf{\Omega}_{IL}|}.$$

De seguida, e com base no caso em que  $M = 3$ , obtiveram-se as expressões de  $\Omega$  e de  $|\Omega|$ , para o caso em que  $M = 4$ , para posteriormente se estabelecerem as expressões correspondentes para o caso em que temos  $M$  equações ( $M > 4$ ).

A partir da expressão (7.2.11) resulta que, para  $M = 4$ ,

$$|\Omega| = |\Omega_{11}| |\Omega_{22 \cdot 1}| |\Omega_{33 \cdot 21}| |\Omega_{44 \cdot 321}|.$$

Para obter o valor de  $|\Omega|$  basta calcular  $|\Omega_{44 \cdot 321}|$ , uma vez que os outros determinantes já foram obtidos no caso em que  $M = 3$ . Assim, partindo da expressão (7.2.12), pode escrever-se que

$$\Omega_{44 \cdot 321} = \Omega_{44 \cdot 32} - \Omega_{41 \cdot 32} \Omega_{11 \cdot 32}^{-1} \Omega_{14 \cdot 32}, \quad (7.2.16)$$

e tendo em conta o resultado (7.2.15), vem

$$\begin{aligned} \Omega_{11 \cdot 32} &= \text{diag} [\sigma_{1t}^2 (1 - \rho_{13t}^2) (1 - \rho_{12 \cdot 3t}^2)], \\ \Omega_{11 \cdot 32}^{-1} &= \text{diag} \left[ \frac{1}{\sigma_{1t}^2 (1 - \rho_{13t}^2) (1 - \rho_{12 \cdot 3t}^2)} \right], \\ \Omega_{44 \cdot 32} &= \text{diag} [\sigma_{4t}^2 (1 - \rho_{43t}^2) (1 - \rho_{42 \cdot 3t}^2)]. \end{aligned}$$

Invocando de novo (7.2.12), tem-se que

$$\Omega_{41 \cdot 23} = \Omega_{41 \cdot 3} - \Omega_{12 \cdot 3} \Omega_{22 \cdot 3}^{-1} \Omega_{42 \cdot 3},$$

onde

$$\begin{aligned} \Omega_{41 \cdot 3} &= \text{diag} \left[ \sigma_{41t} - \frac{\sigma_{13t} \sigma_{43t}}{\sigma_{3t}^2} \right], \quad \Omega_{12 \cdot 3} = \text{diag} \left[ \sigma_{12t} - \frac{\sigma_{13t} \sigma_{23t}}{\sigma_{3t}^2} \right], \\ \Omega_{42 \cdot 3} &= \text{diag} \left[ \sigma_{42t} - \frac{\sigma_{43t} \sigma_{23t}}{\sigma_{3t}^2} \right], \quad \Omega_{22 \cdot 3} = \text{diag} [\sigma_{2t}^2 (1 - \rho_{23t}^2)], \\ \Omega_{22 \cdot 3}^{-1} &= \text{diag} \left[ \frac{1}{\sigma_{2t}^2 (1 - \rho_{23t}^2)} \right]. \end{aligned}$$

Substituindo cada factor pela respectiva expressão, em (7.2.16), vem

$$\begin{aligned} \Omega_{44 \cdot 321} &= \text{diag} [\sigma_{4t}^2 (1 - \rho_{43t}^2) (1 - \rho_{42 \cdot 3t}^2)] \\ &- \text{diag} \left[ \left( \sigma_{41t} - \frac{\sigma_{13t} \sigma_{43t}}{\sigma_{3t}^2} \right) - \frac{\left( \sigma_{12t} - \frac{\sigma_{13t} \sigma_{23t}}{\sigma_{3t}^2} \right) \left( \sigma_{42t} - \frac{\sigma_{43t} \sigma_{23t}}{\sigma_{3t}^2} \right)}{\sigma_{2t}^2 (1 - \rho_{23t}^2)} \right] \\ &\times \text{diag} \left[ \frac{1}{\sigma_{1t}^2 (1 - \rho_{13t}^2) (1 - \rho_{12 \cdot 3t}^2)} \right] \\ &\times \text{diag} \left[ \left( \sigma_{41t} - \frac{\sigma_{13t} \sigma_{43t}}{\sigma_{3t}^2} \right) - \frac{\left( \sigma_{12t} - \frac{\sigma_{13t} \sigma_{23t}}{\sigma_{3t}^2} \right) \left( \sigma_{42t} - \frac{\sigma_{43t} \sigma_{23t}}{\sigma_{3t}^2} \right)}{\sigma_{2t}^2 (1 - \rho_{23t}^2)} \right]. \end{aligned}$$

De seguida, efectuaram-se as possíveis simplificações envolvendo operações com fracções, tendo em conta a relação existente entre a covariância e o coeficiente de correlação e usando (7.2.14).

$$\begin{aligned}
\Omega_{44 \cdot 321} &= \text{diag} [\sigma_{4t}^2 (1 - \rho_{43t}^2) (1 - \rho_{42.3t}^2)] - \text{diag} \left[ \frac{1}{\sigma_{1t}^2 (1 - \rho_{13t}^2) (1 - \rho_{12.3t}^2)} \right] \\
&\times \text{diag} \left\{ \left[ \frac{(\sigma_{3t}^2 \sigma_{41t} - \sigma_{13t} \sigma_{43t}) \sigma_{3t}^2 \sigma_{2t}^2 (1 - \rho_{23t}^2) - (\sigma_{12t} \sigma_{3t}^2 - \sigma_{13t} \sigma_{23t}) (\sigma_{42t} \sigma_{3t}^2 - \sigma_{43t} \sigma_{23t})}{\sigma_{3t}^4 \sigma_{2t}^2 (1 - \rho_{23t}^2)} \right]^2 \right\} \\
&= \text{diag} [\sigma_{4t}^2 (1 - \rho_{43t}^2) (1 - \rho_{42.3t}^2)] \\
&- \text{diag} \left\{ \frac{\sigma_{4t}^2 \sigma_{3t}^8 \sigma_{2t}^4 \sigma_{1t}^2 \left[ \left( \frac{\sigma_{14t}}{\sigma_{1t} \sigma_{4t}} - \frac{\sigma_{13t} \sigma_{43t}}{\sigma_{1t} \sigma_{3t}^2 \sigma_{4t}} \right) (1 - \rho_{23t}^2) - \left( \frac{\sigma_{12t}}{\sigma_{1t} \sigma_{2t}} - \frac{\sigma_{13t} \sigma_{23t}}{\sigma_{1t} \sigma_{3t}^2 \sigma_{2t}} \right) \left( \frac{\sigma_{42t}}{\sigma_{4t} \sigma_{2t}} - \frac{\sigma_{43t} \sigma_{23t}}{\sigma_{4t} \sigma_{3t}^2 \sigma_{2t}} \right) \right]^2}{\sigma_{3t}^8 \sigma_{2t}^4 \sigma_{1t}^2 (1 - \rho_{23t}^2)^2 (1 - \rho_{13t}^2) (1 - \rho_{12.3t}^2)} \right\} \\
&= \text{diag} [\sigma_{4t}^2 (1 - \rho_{43t}^2) (1 - \rho_{42.3t}^2)] \\
&- \text{diag} \left\{ \frac{\sigma_{4t}^2 [(\rho_{14t} - \rho_{13t} \rho_{43t}) (1 - \rho_{23t}^2) - (\rho_{12t} - \rho_{13t} \rho_{23t}) (\rho_{42t} - \rho_{43t} \rho_{23t})]^2}{(1 - \rho_{23t}^2)^2 (1 - \rho_{13t}^2) (1 - \rho_{12.3t}^2)} \right\} \\
&= \text{diag} [\sigma_{4t}^2 (1 - \rho_{43t}^2) (1 - \rho_{42.3t}^2)] \\
&- \text{diag} \left\{ \sigma_{4t}^2 \left[ \frac{(\rho_{14t} - \rho_{13t} \rho_{43t})}{\sqrt{(1 - \rho_{13t}^2) (1 - \rho_{43t}^2)}} - \frac{(\rho_{12t} - \rho_{13t} \rho_{23t}) (\rho_{42t} - \rho_{43t} \rho_{23t})}{(1 - \rho_{23t}^2) \sqrt{(1 - \rho_{13t}^2) (1 - \rho_{43t}^2)}} \right]^2 \right. \\
&\times \left. \frac{(1 - \rho_{23t}^2)^2 (1 - \rho_{13t}^2) (1 - \rho_{43t}^2)}{(1 - \rho_{23t}^2)^2 (1 - \rho_{13t}^2) (1 - \rho_{12.3t}^2)} \right\}.
\end{aligned}$$

Obteve-se depois uma forma mais simplificada, sobre a qual se fizeram mais algumas operações, baseadas na expressão do coeficiente de correlação parcial, referida por (7.2.14), como se descreve de seguida.

$$\begin{aligned}
\Omega_{44 \cdot 321} &= \text{diag} [\sigma_{4t}^2 (1 - \rho_{43t}^2) (1 - \rho_{42.3t}^2)] - \text{diag} \left\{ \sigma_{4t}^2 [\rho_{14.3t} - \rho_{12.3t} \rho_{42.3t}]^2 \frac{(1 - \rho_{43t}^2)}{(1 - \rho_{12.3t}^2)} \right\} \\
&= \text{diag} [\sigma_{4t}^2 (1 - \rho_{43t}^2) (1 - \rho_{42.3t}^2)] \\
&- \text{diag} \left\{ \sigma_{4t}^2 (1 - \rho_{43t}^2) \frac{[\rho_{14.3t} - \rho_{12.3t} \rho_{42.3t}]^2}{(1 - \rho_{12.3t}^2) (1 - \rho_{42.3t}^2)} \frac{(1 - \rho_{12.3t}^2) (1 - \rho_{42.3t}^2)}{(1 - \rho_{12.3t}^2)} \right\} \\
&= \text{diag} [\sigma_{4t}^2 (1 - \rho_{43t}^2) (1 - \rho_{42.3t}^2)] - \text{diag} [\sigma_{4t}^2 (1 - \rho_{43t}^2) (1 - \rho_{42.3t}^2) \rho_{14.32t}^2] \\
&= \text{diag} [\sigma_{4t}^2 (1 - \rho_{43t}^2) (1 - \rho_{42.3t}^2) (1 - \rho_{14.32t}^2)].
\end{aligned}$$

Resultou assim que

$$\Omega_{44 \cdot 321} = \text{diag} [\sigma_{4t}^2 (1 - \rho_{43t}^2) (1 - \rho_{42.3t}^2) (1 - \rho_{14.32t}^2)],$$

e, consequentemente,

$$|\Omega| = |\Omega_{11}| |\Omega_{22}| |\Omega_{33}| |\Omega_{44}| \times \prod_{t=1}^T [(1 - \rho_{12t}^2) (1 - \rho_{32t}^2) (1 - \rho_{31.2t}^2) (1 - \rho_{43t}^2) (1 - \rho_{42.3t}^2) (1 - \rho_{14.32t}^2)].$$

Considerando agora que

$$b_t = (1 - \rho_{12t}^2) (1 - \rho_{32t}^2) (1 - \rho_{31.2t}^2) (1 - \rho_{43t}^2) (1 - \rho_{42.3t}^2) (1 - \rho_{14.32t}^2),$$

vem que  $0 \leq b_t \leq 1$ , uma vez que é o produto de vários factores que tomam valores em  $[0, 1]$ , e também que  $0 \leq \prod_{t=1}^T b_t \leq 1$ . Deste modo, obtém-se

$$|\Omega| \leq |\Omega_{11}| |\Omega_{22}| |\Omega_{33}| |\Omega_{44}|.$$

Supondo que  $|\Omega| \leq |\Omega_{IL}|$  se verifica, por hipótese, para um modelo com  $M$  equações ( $M \in \mathbb{N}$ ), mostra-se de seguida que esse resultado se verifica ainda para um modelo com  $M + 1$  equações. Isto permitiu concluir que  $|\Omega| \leq |\Omega_{IL}|$  se verifica para qualquer número  $M$  de equações, por indução matemática.

Resumindo os resultados obtidos até agora, tem-se

$$M = 2 : \Omega = \Omega_{11} \Omega_{22.1},$$

$$|\Omega| = |\Omega_{11}| |\Omega_{22}| \prod_{t=1}^T (1 - \rho_{12t}^2).$$

$$M = 3 : \Omega = \Omega_{11} \Omega_{22.1} \Omega_{33.21},$$

$$|\Omega| = |\Omega_{11}| |\Omega_{22}| |\Omega_{33}| \prod_{t=1}^T (1 - \rho_{12t}^2) (1 - \rho_{32t}^2) (1 - \rho_{13.2t}^2).$$

$$M = 4 : \Omega = \Omega_{11} \Omega_{22.1} \Omega_{33.21} \Omega_{44.321},$$

$$|\Omega| = |\Omega_{11}| |\Omega_{22}| |\Omega_{33}| |\Omega_{44}| \times \prod_{t=1}^T (1 - \rho_{12t}^2) (1 - \rho_{32t}^2) (1 - \rho_{13.2t}^2) (1 - \rho_{43t}^2) (1 - \rho_{42.3t}^2) (1 - \rho_{14.32t}^2).$$

Admitindo, por hipótese, que se verificam os resultados análogos aos anteriores, para um modelo com  $M$  equações:

$$\Omega = \Omega_{11} \Omega_{22.1} \Omega_{33.21} \Omega_{44.321} \dots \Omega_{MM \cdot (M-1)(M-2) \dots 321},$$

$$\begin{aligned}
|\Omega| &= |\Omega_{11}| |\Omega_{22}| |\Omega_{33}| |\Omega_{44}| \dots |\Omega_{MM}| \\
&\times \prod_{t=1}^T (1 - \rho_{12t}^2) (1 - \rho_{32t}^2) (1 - \rho_{13.2t}^2) (1 - \rho_{43t}^2) (1 - \rho_{42.3t}^2) (1 - \rho_{41.32t}^2) \dots \\
&\times (1 - \rho_{M(M-1)t}^2) (1 - \rho_{M(M-2).(M-1)t}^2) \dots (1 - \rho_{M1.(M-1)(M-2)\dots 32t}^2), \quad (7.2.17)
\end{aligned}$$

vai mostrar-se que se verificam resultados análogos para um modelo com  $(M + 1)$  equações:

$$\Omega = \Omega_{11} \Omega_{22.1} \Omega_{33.21} \dots \Omega_{MM.(M-1)(M-2)\dots 321} \Omega_{(M+1)(M+1).M(M-1)(M-2)\dots 321}, \quad (7.2.18)$$

$$\begin{aligned}
|\Omega| &= |\Omega_{11}| |\Omega_{22}| |\Omega_{33}| |\Omega_{44}| \dots |\Omega_{MM}| |\Omega_{(M+1)(M+1)}| \\
&\prod_{t=1}^T (1 - \rho_{12t}^2) (1 - \rho_{32t}^2) (1 - \rho_{13.2t}^2) (1 - \rho_{43t}^2) (1 - \rho_{42.3t}^2) (1 - \rho_{41.32t}^2) \dots \\
&\times (1 - \rho_{M(M-1)t}^2) (1 - \rho_{M(M-2).(M-1)t}^2) \dots (1 - \rho_{M1.(M-1)(M-2)\dots 32t}^2) \\
&\times (1 - \rho_{(M+1)Mt}^2) (1 - \rho_{(M+1)(M-1).Mt}^2) \dots (1 - \rho_{(M+1)1.M(M-1)(M-2)\dots 32t}^2).
\end{aligned}$$

Para chegar a esse resultado, calculou-se o determinante de  $\Omega$  a partir da expressão (7.2.18). Tendo em conta que se pode substituir os primeiros factores, de acordo com a expressão (7.2.17), admitida por hipótese, e mantendo apenas o último determinante, que é desconhecido, vem

$$\begin{aligned}
|\Omega| &= |\Omega_{11}| |\Omega_{22}| |\Omega_{33}| |\Omega_{44}| \dots |\Omega_{MM}| \\
&\prod_{t=1}^T (1 - \rho_{12t}^2) (1 - \rho_{32t}^2) (1 - \rho_{13.2t}^2) (1 - \rho_{43t}^2) (1 - \rho_{42.3t}^2) (1 - \rho_{41.32t}^2) \times \dots \\
&\times (1 - \rho_{M(M-1)t}^2) (1 - \rho_{M(M-2).(M-1)t}^2) \dots (1 - \rho_{M1.(M-1)(M-2)\dots 32t}^2) \\
&\times |\Omega_{(M+1)(M+1).M(M-1)(M-2)\dots 321}|.
\end{aligned}$$

Para obter o determinante  $|\Omega_{(M+1)(M+1).M(M-1)(M-2)\dots 321}|$ , considerou-se que

$$\begin{aligned}
\Omega_{22.1} &= \text{diag} [\sigma_{2t}^2 (1 - \rho_{21t}^2)], \\
\Omega_{33.21} &= \text{diag} [\sigma_{3t}^2 (1 - \rho_{32t}^2) (1 - \rho_{31.2t}^2)], \\
\Omega_{44.321} &= \text{diag} [\sigma_{4t}^2 (1 - \rho_{43t}^2) (1 - \rho_{42.3t}^2) (1 - \rho_{41.32t}^2)],
\end{aligned}$$

donde resulta que

$$\Omega_{MM.(M-1)(M-2)\dots 321} = \text{diag} [\sigma_{M,t}^2 (1 - \rho_{M(M-1)t}^2) \dots (1 - \rho_{1M.(M-1)(M-2)\dots 32t}^2)],$$

e ainda que

$$\begin{aligned} \mathbf{\Omega}_{(M+1)(M+1) \cdot M(M-1) \dots 321} &= \\ &= \text{diag} \left[ \sigma_{(M+1)t}^2 \left( 1 - \rho_{(M+1)Mt}^2 \right) \left( 1 - \rho_{(M+1)(M-1)Mt}^2 \right) \dots \left( 1 - \rho_{(M+1)1.M(M-1)(M-2) \dots 32t}^2 \right) \right]. \end{aligned} \quad (7.2.19)$$

Por outro lado,

$$\begin{aligned} \mathbf{\Omega}_{(M+1)(M+1) \cdot M(M-1) \dots 321} &= \mathbf{\Omega}_{(M+1)(M+1) \cdot M(M-1) \dots 32} - \mathbf{\Omega}_{(M+1)1 \cdot M(M-1) \dots 32} \mathbf{\Omega}_{11 \cdot M(M-1) \dots 32}^{-1} \\ &\quad \times \mathbf{\Omega}_{1(M+1) \cdot M(M-1) \dots 32}, \end{aligned}$$

e cada uma destas matrizes pode ser desenvolvida de acordo com a expressão (7.2.19), obtendo-se o seguinte,

$$\begin{aligned} \left| \mathbf{\Omega}_{(M+1)(M+1) \cdot M(M-1) \dots 321} \right| &= \\ &= \left| \text{diag} \left[ \sigma_{(M+1)t}^2 \left( 1 - \rho_{(M+1)Mt}^2 \right) \left( 1 - \rho_{(M+1)(M-1)Mt}^2 \right) \dots \left( 1 - \rho_{(M+1)1.M(M-1) \dots 32t}^2 \right) \right] \right|. \end{aligned}$$

Consequentemente,

$$\begin{aligned} |\mathbf{\Omega}| &= |\mathbf{\Omega}_{11}| |\mathbf{\Omega}_{22}| |\mathbf{\Omega}_{33}| |\mathbf{\Omega}_{44}| \dots |\mathbf{\Omega}_{MM}| |\mathbf{\Omega}_{(M+1)(M+1)}| \\ &\quad \times \prod_{t=1}^T \left( 1 - \rho_{12t}^2 \right) \left( 1 - \rho_{32t}^2 \right) \left( 1 - \rho_{13.2t}^2 \right) \left( 1 - \rho_{43t}^2 \right) \left( 1 - \rho_{42.3t}^2 \right) \left( 1 - \rho_{41.32t}^2 \right) \dots \\ &\quad \times \left( 1 - \rho_{M(M-1)t}^2 \right) \left( 1 - \rho_{M(M-2).(M-1)t}^2 \right) \dots \left( 1 - \rho_{M1.(M-1)(M-2) \dots 32t}^2 \right) \\ &\quad \times \left( 1 - \rho_{(M+1)Mt}^2 \right) \left( 1 - \rho_{(M+1)(M-1)Mt}^2 \right) \dots \left( 1 - \rho_{(M+1)1.M(M-1) \dots 32t}^2 \right). \end{aligned}$$

Se considerarmos

$$\begin{aligned} c_t &= \left( 1 - \rho_{12t}^2 \right) \left( 1 - \rho_{32t}^2 \right) \left( 1 - \rho_{13.2t}^2 \right) \left( 1 - \rho_{43t}^2 \right) \left( 1 - \rho_{42.3t}^2 \right) \left( 1 - \rho_{41.32t}^2 \right) \dots \\ &\quad \times \left( 1 - \rho_{M(M-1)t}^2 \right) \left( 1 - \rho_{M(M-2).(M-1)t}^2 \right) \dots \left( 1 - \rho_{M1.(M-1)(M-2) \dots 32t}^2 \right) \\ &\quad \times \left( 1 - \rho_{(M+1)Mt}^2 \right) \left( 1 - \rho_{(M+1)(M-1)Mt}^2 \right) \dots \left( 1 - \rho_{(M+1)1.M(M-1) \dots 32t}^2 \right), \end{aligned}$$

vem que  $0 \leq c_t \leq 1$ , uma vez que é o produto de factores que variam todos em  $[0, 1]$ , pelo que  $0 \leq \prod_{t=1}^T c_t \leq 1$ . Deste modo, obtém-se que

$$|\mathbf{\Omega}| \leq \underbrace{|\mathbf{\Omega}_{11}| |\mathbf{\Omega}_{22}| |\mathbf{\Omega}_{33}| |\mathbf{\Omega}_{44}| \dots |\mathbf{\Omega}_{MM}|}_{|\mathbf{\Omega}_{IL}|} |\mathbf{\Omega}_{(M+1)(M+1)}|.$$

Provou-se assim que, qualquer que seja o número  $M$  de equações do modelo, se verifica

$$|\mathbf{\Omega}| \leq |\mathbf{\Omega}_{IL}|.$$

Donde se conclui, com base no Critério.2, que o estimador  $GMMIC$  é assintoticamente mais eficiente do que o estimador  $GMMIL$ . □

# Referências Bibliográficas

Amemiya, T. (1982). "Two stage least absolute deviation estimators", *Econometrica* **50**: 689-711.

Andrade, J. (1993). *Análise econométrica: uma introdução para economistas e gestores*, Texto Editora, Lisboa.

Arellano, M. (2003). *Panel data econometrics*, Oxford University Press, New York.

Baltagi, B. (2008). *Econometrics, fourth edition*, Springer-Verlag, Berlin Heidelberg.

Bilodeau, M. e Duchesne, P. (2000). "Robust estimation of the SUR model", *La Review Canadienne de Statistique* **28**: 277-288.

Buchinsky, M. (1997). "Recent advances in quantile regression models: a practical guideline for empirical research", *Journal of Human Resources*, **33**: 88-126.

Casella, G. e Berger, R. (1990). *Statistical inference*, Duxbury Press, Belmont.

Chen, L. e Portnoy, S. (1996). "Two-stage regression quantiles and two-stage trimmed least squares estimators for structural equation models", *Communications Statistics - Theory and Methods* **25**:1005-1032.

Chen, L., Thompson, P. e Hung, H. (2000a). "The symmetric type two-stage trimmed least squares estimator for the simultaneous equations model", *Statistica Sinica* **10**: 1243-1255.

Chen, L., Thompson, P. e Chuang, H. (2000b). "Mallow's type bounded influence regression quantile for linear regression model and simultaneous equations model", *Sankya: The Indian Journal of Statistics* **62**: 217-232.

Chen, L., Liang, K. e Liu, C. (2001). "Two-stage Welsh's trimmed mean for simultaneous equations model", *Australian New-Zeeland Journal of Statistics* **43**: 481-492.

Christ, C. (1966). *Econometric models and methods*, Wiley, New York.



Cochrane, J. (2001). *Asset pricing, revised edition*, Princeton University Press, Princeton and Oxford.

Desai, M. (1976). *Applied econometrics*, Philip Allen, Oxford.

Draper, N. e Smith, H. (1998). *Applied regression analysis, third edition*, Wiley, New York.

Donald, S., Imbens, G. e Newey, W. (2009). "Choosing instrumental variables in conditional moment restriction models", *Journal of Econometrics* **152**: 28-36.

Engle, R., Hendry, D. e Richard, J. (1983). "Exogeneity", *Econometrica* **51**: 277-304.

Fraiman, R., Meloche, J., García-Escudero, L., Gordaliza, A., He, X., Maronna, R., Yohai, V., Sheather, S., McKean, J., Small, C. e Wood, A. (1999). "Multivariate L-estimation", *Test* **8**: 255-317.

Ghysels, E. e Hall, A. (2002). "Editor's introduction to JBES twentieth anniversary issue on generalized method of moments", *Journal of Business & Economic Statistics* **20**: 441-441.

Gnanadesikan, R. e Kettenring, J. (1972). "Robust estimates, residuals and outlier detection with multiresponse data", *Biometrics* **28**: 81-124.

Goldberger, A. (1964). *Economic theory*, John Wiley & Sons, New York.

Greene, W. (2003). *Econometric analysis, fifth edition*, Prentice-Hall, New Jersey.

Gujarati, D. (2003). *Basic econometrics, fourth edition*, McGraw-Hill, New York.

Haavelmo, T. (1944). "The probability approach in econometrics", *Journal of Econometrics* **100**: 3-5.

Hall, A. (1993). "Some aspects of generalized method of moments estimation", *Handbook of Statistics* **11**: 393-417.

Hall, A. (2001). "Generalized method of moments", in *Companion in Theoretical Econometrics*, B. Baltagi, Basil Blackwell, Oxford: 230-255.

Hall, A. (2005). *Generalized Method of Moments*, Oxford University Press, New York.

Hamilton, J. (1994). *Time series analysis*, Princeton University Press, New Jersey.

Hampel, F., Ronchetti, E., Rousseeuw, P. e Stahel, W. (1986). *Robust statistics: The approach based on influence functions*, John Wiley & Sons, New York.

- Hansen, B. e West, K. (2002). "Generalized method of moments and macroeconomics", *Journal of Business & Economic Statistics* **20**: 460-469.
- Hansen, L. (1982). "Large sample properties of generalized method of moments estimators", *Econometrica* **50**: 1029-1054.
- Hausman, J. (1983). "Specification and estimation of simultaneous equation models", In Z. Griliches, M. D. Intriligator, *Handbook of Econometrics*, North-Holland Publishing Company.
- Hendry, D. (1980). "Econometrics – alchemy or science?", *Economica* **47**: 387-406.
- Hodges, J. e Lehman, E. (1963). "Estimate of location based on rank tests", *Annals of Mathematical Statistics* **34**: 598-611.
- Huber, P. (1964). "Robust estimation of a location parameter", *Annals of Mathematical Statistics* **35**: 73-101.
- Huber, P. (1973). "Robust regression: Asymptotics conjectures and monte carlo", *Annals of Mathematical Statistics* **1**: 729-821.
- Imbens, G. (2002). "Generalized method of moments and empirical likelihood", *Journal of Business & Economic Statistics* **20**: 493-503.
- Imbens, G. e Spady, R. (2002). "Confidence intervals in generalized method of moments models", *Journal of Econometrics* **107**: 87-98.
- Intriligator, M., Bodkin, R. e Hsiao, C. (1996). *Econometric models, techniques, and applications, second edition*, Prentice Hall, Upper Saddle River, New Jersey.
- Jagannathan, R., Skoulakis, G. e Wang, Z. (2002). "Generalized method of moments: applications in finance", *Journal of Business & Economic Statistics* **20**: 470-481.
- Johnston, J. e DiNardo, J. (1997), *Econometric methods, fourth edition*, MacGraw-Hill, New York.
- Judge, G., Griffiths, W., Lutkepohl, Hill, R. e Lee, T. (1988). *Introduction to the theory and practice of econometrics, second edition*, John Wiley & Sons, New York.
- Kapteyn, A. e Fiebig, D. (1981). "When are two-stage and three-stage least squares estimation identical?", *Economic Letters* **8**: 53-57.
- Kim, T. e Muller, K. (2004). "Two-stage quantile regression when the first stage is based on quantile regression", *Econometrics Journal* **7**: 218-231.

Kim, T. e Muller, C. (2007). "Two-stage Huber estimation", *Journal of Statistical Planning and Inference* **137**: 405-418

Klein, L. (1950). *Economic fluctuations in the United States 1921-1941*. Cowles Commission Monograph. **11**, John Wiley & Sons, New York.

Kmenta, J. (1997). *Elements of econometrics, second edition*, The University of Michigan Press, Ann Arbor.

Koenker, R. e Basset, G. (1978). "Regression quantiles", *Econometrica* **46**: 33-50.

Koenker, R. e Portnoy, S. (1990). "M estimation of multivariate regressions", *Journal of the American Statistical Association* **85**: 1060-1068.

Krasker, W. (1986). "Two-stage bounded-influence estimators for simultaneous equations models", *Journal of Business and Economic Statistics* **4**: 437-444.

Krasker, W. e Welsch, R. (1985). "Resistant estimation for simultaneous- equations models using weighted instrumental variables", *Econometrica* **53**: 1475-1488.

Krishnakumar, J. (2007). "Challenges facing econometrics in the 21st century", *Contemporary issues and ideas in social sciences*, **3**.

Krishnakumar, J. e Ronchetti, E. (1997). "Robust estimators for simultaneous equations models", *Journal of Econometrics*, **78**: 295-314.

Leamer, E. (1983). "Lets take the con out of econometrics", *American Economic Review* **73**: 31-43.

Leamer, E. (1985). "Vector autoregressive for causal inference", In Brunner and A. Meltzer (eds), *Understanding Monetary Regimes, Supplement to Journal of Monetary Economics*, 255-304.

Lopuhaä, H. (1991). "Multivariate tau-estimators for location and scatter", *The Canadian Journal of Statistics* **19**: 307:321.

Machado, J. e Santos Silva, J. (2002). "50 anos de ensino da Econometria em Portugal", *Economia* **26**: 95-112.

Magalhães, M. (1994). "Métodos de previsão para gestão: identificação e estimação de modelos de equações simultâneas", *Temas em Métodos Quantitativos para Gestão*, **14**, Giesta, ISCTE.

Mardia, K., Kent, J. e Bibby, J. (1979). *Multivariate analysis*, Academic Press, New York.

Maronna, R. e Yohai, V. (1997). "Robust estimation in simultaneous equations models", *Journal of Statistical Planning and Inference* **57**: 233-244.

Maronna, R. e Zamar, R. (2002). "Robust estimates of location and dispersion for high-dimensional datasets", *Technometrics* **44**: 307-317.

Maronna, R., Martin, R. e Yohai, V. (2006). *Robust statistics, theory and methods*, John Wiley & Sons, Chichester.

Mittelhammer, R., Judge, G. e Miller, D. (2000). *Econometric foundations*, Cambridge University Press, Cambridge.

Newey, W. e McFadden, D. (1994). "Large sample estimation and hypothesis testing", *Handbook of Econometrics*, R. Engle and D. L. McFadden, Elsevier Science, Amsterdam: 2113-2247.

Peña, D. e Prieto, F. (2001). "Multivariate outlier detection and robust covariance matrix estimation", *Technometrics* **43**: 286-300.

Penalva, H. (2003). *Modelo Linear de Equações Simultâneas*, tese de Mestrado, Faculdade de Ciências, Universidade de Lisboa, Lisboa.

Pinheiro, A. e Galego, M. (1999). *Econometria*, Manuais da Universidade de Évora, Évora.

Pires, A. e Branco, J. (2007). *Introdução aos métodos estatísticos robustos*, Sociedade Portuguesa de Estatística, Lisboa.

Prucha, I. e Kelejian, H. (1984). "The structure of simultaneous equation estimators: A generalization towards nonnormal disturbances", *Econometrica* **52**: 721-736.

R Development Core Team (2007). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.

Ronchetti, E. e Trojani, F. (2001). "Robust inference with GMM estimators", *Journal of Econometrics* **101**: 37-69.

Rousseeuw, P. (1984). "Least median of squares regression", *Journal of the American Statistical Association* **79**: 871-880.

Rousseeuw, P. (1985). "Multivariate estimation with high breakdown point", *Mathematical Statistics and Applications*, W. Grossman, G. Pflug, I. Vincze and W. Wertz, Reidel Publishing Company, Dordrecht (co-published with Akadémiai Kiadó Budapest): 283-297.

Rousseeuw, P. e Yohai, V. (1984). "Robust regression by means of S-estimators", in *Robust and Nonlinear Time Series Analysis*, J. Franke, W. Härdle and D. Martin, Lecture Notes in Statistics, **26**, Springer Verlag, Berlin: 256-272.

Rousseeuw, P. e Leroy, A. (1987). *Robust regression and outlier detection*, John Wiley & Sons, New York.

Rousseeuw, P. e Croux, C. (1993). "Alternatives to median absolute deviation", *Journal of the American Statistical Association* **88**: 1273-1283.

Rousseeuw, P. e Croux, C. (1994). "The bias of k-step M-estimators", *Statistics and Probability Letters* **20**: 411-420.

Rousseeuw, P. e Van Driessen, K. (1999). "A fast algorithm for the minimum covariance determinant estimator", *Technometrics* **41**: 212-223.

Ryan, T. (1997). *Modern regression methods*, John Wiley & Sons, New York.

Sakata, S. (2007). "Instrumental variable estimation based on conditional median restriction", *Journal of Econometrics* **141**: 350-282.

Schmidt, P. (1976) *Econometrics*, Marcel Dekker, New York.

Seber, J. (1984). *Multivariate Observations*, John Wiley & Sons, New York.

Singleton, K. (2006). *Empirical dynamic asset-pricing: model specification and econometric assessment*, Princeton University Press.

Srivastava, V. e Tiwari, R. (1978). "Efficiency of two-stage and three-stage least squares estimators", *Econometrica* **46**: 1495-1498.

Staudte, R. e Sheather, S. (1990). *Robust estimation and testing*, John Wiley & Sons, New York.

Theil, H. (1971). *Principles of econometrics*, Wiley, New York.

Tintner, G. (1953). "The definition of econometrics", *Econometrica* **21**: 31-40.

Welsh, A. (1987). "The trimmed mean in linear model", *Annals of Statistics* **15**: 20-36.

White, H. (1980). "A heteroskedasticity-consistent covariance matrix estimator and a direct test for a heteroskedasticity", *Econometrica* **48**: 817-838.

Wooldridge, J. (2009). *Introductory Econometrics: A Modern Approach, fourth edition*, South-Western College Pub.

Yohai, V. (1987). "High breakdown-point and high efficiency robust estimates for

regression", *Annals of Statistics* **15**: 642-656.

Yohai, V. e Zamar, R. (1988). "High breakdown-point estimates of regression by means of minimization of a efficient scale", *Journal of the American Statistical Association* **83**: 406-413.

Zellner, A. e Theil, H. (1962). "Three-stege least squares: Simultaneous estimation of simultaneous equations", *Econometrica* **30**: 54-57.